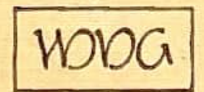
KARL SEILER

PHYSIK UND TECHNIK DER HALBLEITER



WISSENSCHAFTLICHE VERLAGSGESELLSCHAFT M.B.H.
STUTTGART

KARL SEILER

PHYSIK UND TECHNIK DER HALBLEITER

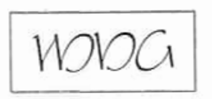
PHYSIK UND TECHNIK

HERAUSGEGEBEN VON DR. FRITZ GÖSSLER

BAND 7

PROF. DR. KARL SEILER

PHYSIK UND TECHNIK DER HALBLEITER



WISSENSCHAFTLICHE VERLAGSGESELLSCHAFT M.B.H.

STUTTGART

1964

PHYSIK UND TECHNIK DER HALBLEITER

Von

PROF. DR. KARL SEILER
Freiburg i. Br.

unter Mitarbeit von DR. HANS P. KLEINKNECHT
Zürich

Mit 127 Abbildungen und 18 Tafeln

WISSENSCHAFTLICHE VERLAGSGESELLSCHAFT M.B.H.
STUTTGART
1964



Satz und Druck: Buchdruckerei Ungeheuer & Ulmer Ludwigsburg

Meiner lieben Frau

und allen, die mir beim Aufbau der Firma INTERMETALL geholfen haben,

gewidmet

Dr. J. BITTNER	W. BOOS	W. BREUKER
Dr. G. EICHHOFF	G. GÄRTLEIN	A. GERLACH
P. GÖTTINGER	H. GÜNDEL	E. HAUGG
Dr. H. HINRICHS	H. KNABE	Dr. H. KNAU
R. KNOCH	Dr. G. LORENZ	H. PAPENHUSEN †
G. PFÜTZNER	Dr. W. SANDERMANN	Dr. D. SAUTTER
R. STASEK	Dr. G. STUTZER	R. THIEMANN



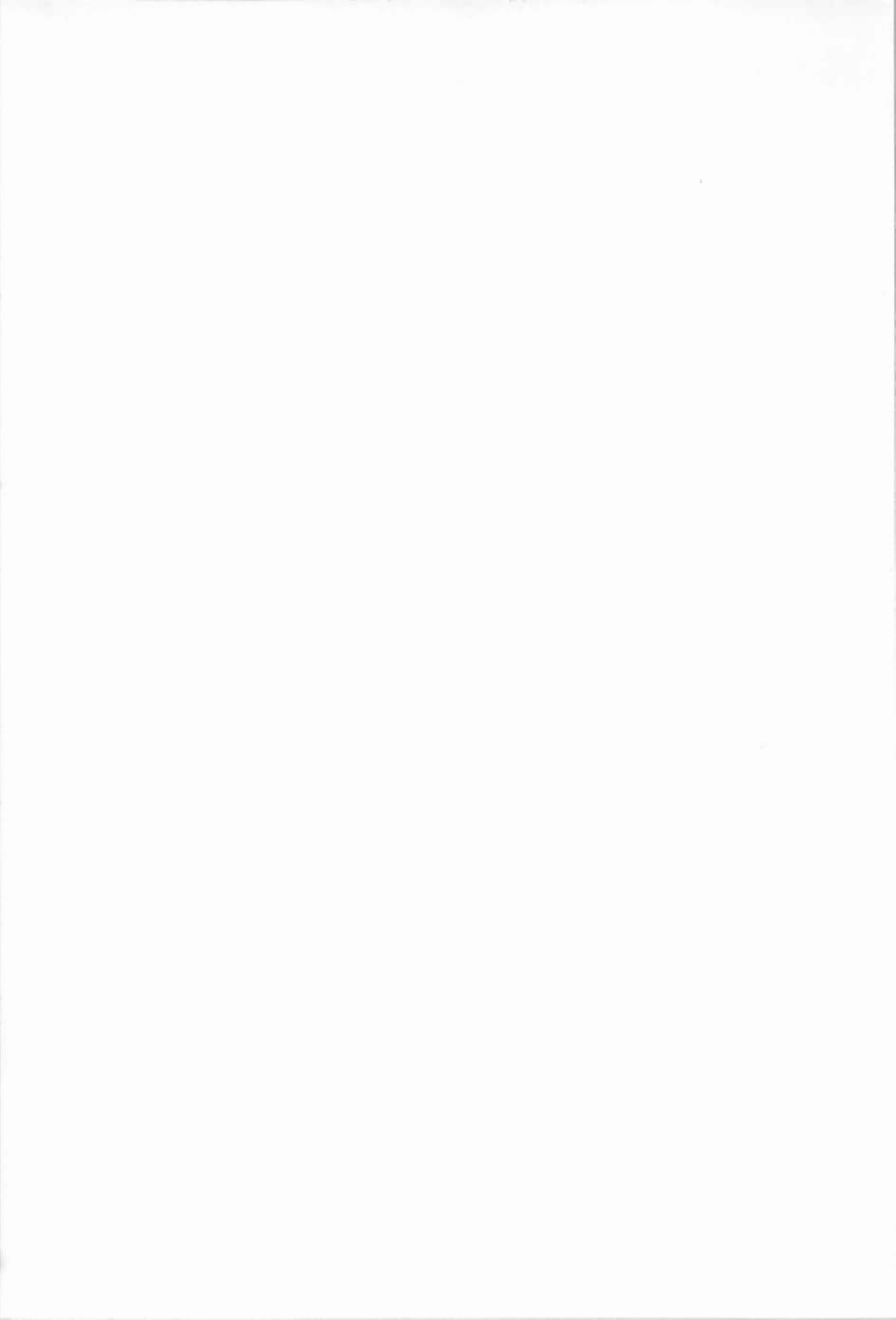
Vorwort

Das vorliegende Buch enthält im wesentlichen Vorlesungen, die ich seit mehr als zehn Jahren an der Technischen Hochschule Stuttgart halte. Diese Vorlesungen haben zum Ziel, Studenten der Elektrotechnik und Physik in möglichst anschaulicher Form in die physikalische Wirkungsweise des *PN*-Übergangs und des Transistors einzuführen. Diese Form ist in den ersten fünf Kapiteln beibehalten worden. Im 6. Kapitel über den Transistor wurden in Abweichung von der Vorlesung neben eingestreuten anschaulichen Überlegungen die vielfältigen Einflüsse auf die Transistorparameter quantitativ formuliert. Nur auf diese Weise kann der in der Praxis von der Anwendung her bestimmte Kompromiß gefunden werden. Bei der Erörterung der Grundlagen wurde ein Kapitel über die wellenmechanische Behandlung des Halbleiterkristalls eingefügt. Diese Darstellung ist sehr gedrängt und auch nicht mehr als eine Einführung.

Wesentlichen Anteil an der Gestaltung der Schrift hat Herr Dr. H. P. Kleinknecht. Er hat nicht nur das Vorlesungsmanuskript druckfertig gemacht, sondern auch an vielen Stellen abgerundet und ergänzt. Das 3. Kapitel stammt von ihm und außerdem viele Formulierungen im 6. Kapitel. Besonderen Dank schulde ich Herrn Diplomphysiker W. Neft für Kritik und die mühsame Korrekturarbeit. Der Wissenschaftlichen Verlagsgesellschaft habe ich für eine angenehme Zusammenarbeit und die gute Ausstattung des Buches zu danken.

Daß ich während der Aufbauphase einer Firma überhaupt Zeit hatte, Vorlesungen zu halten und etwas Forschung zu treiben, verdanke ich meinen ausgezeichneten Mitarbeitern. Deshalb ist ihnen diese Arbeit gewidmet.

Freiburg i. Br., im August 1963 K. Seiler



Inhaltsverzeichnis

Sechstes Kapitel

Verzeichnis der wichtigsten Formelzeichen

A — Austauschintegral, Fläche des PN-Übergangs
A_I — Gleichstromverstärkung für inverse Basisschaltung
A_N — Gleichstromverstärkung für normale Basisschaltung
A_O — Transportfaktor der Stromverstärkung
a — Gitterkonstante, halbe Plattendicke (Abb. 4.3), Emitter-Radius (Abb. 6.10b)
a_1 — Emitter-Basis-Abstand (Abb. 6.25)
a_2 — Emitterbreite (Abb. 6.25)
a_3 — Länge der Emitter- und Basisflecken (Abb. 6.25)
a_{11} — siehe Gl. 6.19
a_{12} — siehe Gl. 6.20
a_{21} — siehe Gl. 6.20
a_{22} — siehe Gl. 6.21
$\tilde{a}_{11}$ — Gl. 6.19 mit $\tanh(W/L_p) \to (\tanh(\eta W/L_p)$
$\tilde{a}_{21}$ — Gl. 6.20 mit $\sinh(W/L_p) \to (\sinh(\eta W/L_p)$
α — siehe Gl. 3.36, Wechselstromverstärkung in Basisschaltung
α_0 — Wechselstromverstärkung in Basisschaltung bei niedriger Frequenz
α' — Wechselstromverstärkung des inneren Transistors in normaler Basisschaltung
α'_I — Wechselstromverstärkung des inneren Transistors in inverser Basisschaltung

B — magnetische Induktion
b — siehe Gl. 4.85, Störstellengradient im graduellen PN-Übergang, Kollektorradius (Abb. 6.10b)
b^* — Breite der Ringfläche, innerhalb deren die Oberflächenrekombination erfolgt
β — Wechselstromverstärkung in Emitterschaltung

C — Coulomb-Integral, Kapazität
C_D — Diffusionskapazität pro Flächeneinheit
C_{DE} — Emitterdiffusionskapazität
C_F — Feldzonen- oder Raumladungs-Kapazität pro Flächeneinheit
C_{FC} — Kollektorraumladungskapazität
C_{FCA} — äußere Kollektorraumladungskapazität
C_{FCI} — innere Kollektorraumladungskapazität
C_{FE} — Emitterraumladungskapazität
C_{pD} — Löcheranteil der Diffusionskapazität pro Flächeneinheit
c — Innenradius des Basiskontaktes (Abb. 6.10b)

D_n — Diffusionskonstante der Elektronen
D_p — Diffusionskonstante der Löcher
D^* — ambipolare Diffusionskonstante (Gl. 5.42)
Δ — $\text{div grad} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$
Δ'_y — Gl. 6.78

E Energie
E_A Aktivierungsenergie der Akzeptoren
E_{At} atomarer Energieeigenwert
E_C Energie des unteren Randes des Leitungsbandes
E_D Aktivierungsenergie der Donatoren
E_F FERMI-Energie
E_G Bandabstand
E_0 Energie am Minimum der Energiefläche im k-Raum
E_V Energie des oberen Randes des Valenzbandes
e Elementarladung, Basis des natürlichen Logarithmus
$\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ Einheitsvektoren
ε relative Dielektrizitätskonstante
ε_0 absolute Dielektrizitätskonstante des Vakuums

F elektrische Feldstärke, Fläche des PN-Übergangs
F_1 Massenwirkungskonstante Gl. 4.1
F_2 Massenwirkungskonstante Gl. 4.9
F_3 Massenwirkungskonstante Gl. 4.10
F_{av} Kritische Feldstärke für Avalanche-Durchbruch
F_C Wirksame Kollektorfläche
F_E Fläche des Emitterübergangs
$F_{\max}$ maximale Feldstärke im PN-Übergang
f Frequenz
f_E Emittergrenzfrequenz
f_L Laufzeitgrenzfrequenz
f_{LB} Basis-Laufzeitgrenzfrequenz
f_{LC} Kollektor-Laufzeitgrenzfrequenz
$f_{\max}$ maximale Oszillationsfrequenz
f_α α-Grenzfrequenz
f_β β-Grenzfrequenz
$f_{\beta 1}$ Frequenz, bei der $\beta = 1$ ist
$f(E)$ FERMI-Funktion
$f_D(E)$ FERMI-Funktion für die Besetzung von Donatoren
$f_p(E)$ FERMI-Funktion für Löcher
φ elektrisches Potential

G Länge des Grundgebietes
G_D Betrag des (negativen) Leitwerts der Tunneldiode
G_G Leitwert des Generators
G_L Leitwert der Last
G_p Löcheranteil des Wirkleitwerts des PN-Überganges pro Flächeneinheit
g Generationsrate
g_E Realteil des Eingangsleitwerts
g_G Realteil des Generatorleitwerts
g_L Generationsrate durch Lichtanregung, Realteil der Lastadmittanz
γ Emitterwirkungsgrad
$\tilde{\gamma}$ gleich γ aus Gl. 6.25, wenn $\tanh\left(\frac{W}{L_p}\right)$ durch $\tanh\left(\frac{\eta W}{L_p}\right)$ ersetzt wird

h PLANCKsche Konstante
$\hbar$ $\equiv h/2\pi$

h_{11B}, h_{12B}, h_{21B}, h_{22B}	Vierpolelemente der Hybrid-Matrix (Gl. 6.126) in Basisschaltung
h_{11E}, h_{12E}, h_{21E}, h_{22E}	Vierpolelemente der Hybrid-Matrix (Gl. 6.126a) in Emitterschaltung
η	$\equiv \sqrt{1 + i\,\omega\,\tau_p}$
I	Elektrischer Strom
I_B	Basisstrom
I_C	Kollektorstrom
I_{CBO}	Kollektor-Basis-Strom bei offenem Emitter
I_{Cn}	Elektronenanteil des Kollektorstroms
I_{Cp}	Löcheranteil des Kollektorstroms
I_C^*	siehe Gl. 6.37
I_E	Emitterstrom
I_{En}	Elektronenanteil des Emitterstroms
I_{Ep}	Löcheranteil des Emitterstroms
I_E^*	siehe Gl. 6.36
I_F	Flußstrom
I_{max}	Maximal zulässiger Strom
I_{pO}	Oberflächenrekombinationsstrom der Löcher
I_{pV}	Volumenrekombinationsstrom der Löcher
I_s	Sättigungsstrom
i	$\equiv \sqrt{-1}$
i_B	Basiswechselstrom
i_C	Kollektorwechselstrom
i_{Cn}	Elektronenanteil des Kollektorwechselstroms
i_{Cp}	Löcheranteil des Kollektorwechselstroms
i_D	siehe Abb. 5.54
i_E	siehe Abb. 5.54, Emitterwechselstrom
i_{En}	Elektronenanteil des Emitterwechselstroms
i_{Ep}	Löcheranteil des Emitterwechselstroms
i_G	siehe Abb. 5.54
i_L	siehe Abb. 5.54
j	Stromdichte
j_D	Diffusionsstromdichte
j_D'	Diffusionsteilchenstromdichte
j_F	Feldstromdichte
j_K	Kurzschlußstromdichte
j_L	Photostromdichte
j_n	Elektronenstromdichte
j_{ns}	Elektronenanteil der Sättigungsstromdichte
j_p	Löcherstromdichte
j_{ps}	Löcheranteil der Sättigungsstromdichte
j_s	Sättigungsstromdichte
K	Kraft, Wärmewiderstand
k	Boltzmannkonstante, Wellenvektor bzw. -zahl
k_0	Wellenzahl am Minimum der Energiefläche im k-Raum

L Induktivität
L_n Diffusionslänge der Elektronen
L_p Diffusionslänge der Löcher
L_{pW} Eindringtiefe der Löcherwelle
l Nummer der Nachbaratome, Dicke der Raumladungszone
l_C Dicke der Kollektorsperrschicht
l_E Dicke der Emittersperrschicht
l_0 Dicke der Raumladungszone bei $U = 0$
λ Wellenlänge

m freie Elektronenmasse
m_{eff} effektive Masse
m_l longitudinale effektive Elektronenmasse
m_n effektive Elektronenmasse
m_p effektive Löchermasse
m_t transversale effektive Elektronenmasse
m_1 effektive Masse der schweren Löcher
m_2 effektive Masse der leichten Löcher
μ Beweglichkeit
μ_n Elektronenbeweglichkeit
μ_p Löcherbeweglichkeit

N Zahl der Atome im Grundgebiet
N_A Akzeptorendichte
N_{A^+} Dichte der neutralen Akzeptoren
N_{A^-} Dichte der ionisierten Akzeptoren
N_{A^C} Akzeptorendichte am Kollektorübergang
N_C effektive Zustandsdichte des Leitungsbandes
N_D Donatorendichte
$N_{D^\times}$ Dichte der neutralen Donatoren
N_{D^+} Dichte der ionisierten Donatoren
N_{D^C} Donatorendichte in der Basis am Kollektorrand
N_{D^E} Donatorendichte in der Basis am Emitterrand
N_{D^O} Donatorendichte an der Oberfläche
$N_{G\max}$ vom Generator maximal verfügbare Leistung
N_V effektive Zustandsdichte des Valenzbandes
N_{verl} maximal zulässige Verlustleistung
N_W Wechselstromausgangsleistung
$N(E)$ Zustandsdichte (pro Energie-Intervall)
n Elektronendichte, ganze Zahl
n_i Eigenleitungsdichte
n_N Gleichgewichtsdichte der Elektronen im N-Gebiet
n_P Gleichgewichtsdichte der Elektronen im P-Gebiet
n_{PC} Gleichgewichtsdichte der Elektronen im Kollektor
n_{PE} Gleichgewichtsdichte der Elektronen im Emitter
ν Frequenz
∇ $\text{grad} \equiv \mathbf{e}_x \frac{\partial}{\partial x} + \mathbf{e}_y \frac{\partial}{\partial y} + \mathbf{e}_z \frac{\partial}{\partial z}$

∇_k Gradient im k-Raum

ω Kreisfrequenz $\equiv 2\pi f$

P Impuls
p Löcherdichte

p_g Gleichstromanteil der Löcherdichte
p_N Gleichgewichtsdichte der Löcher im N-Gebiet
p_P Gleichgewichtsdichte der Löcher im P-Gebiet
p_{PC} Gleichgewichtsdichte der Löcher im Kollektor
p_{PE} Gleichgewichtsdichte der Löcher im Emitter
$\tilde{p}$ Wechselstromanteil der Löcherdichte
Ψ zeitabhängige SCHRÖDINGER-Funktion
$\overline{\Psi}$ Wellenfunktion eines Wellenpakets
ψ zeitunabhängige SCHRÖDINGER-Funktion
ψ_{At} atomare Eigenfunktion

Q elektrische Ladung
Q_B Ladung der Löcher in der Basis
q Ladung eines Leitfähigkeitsträgers

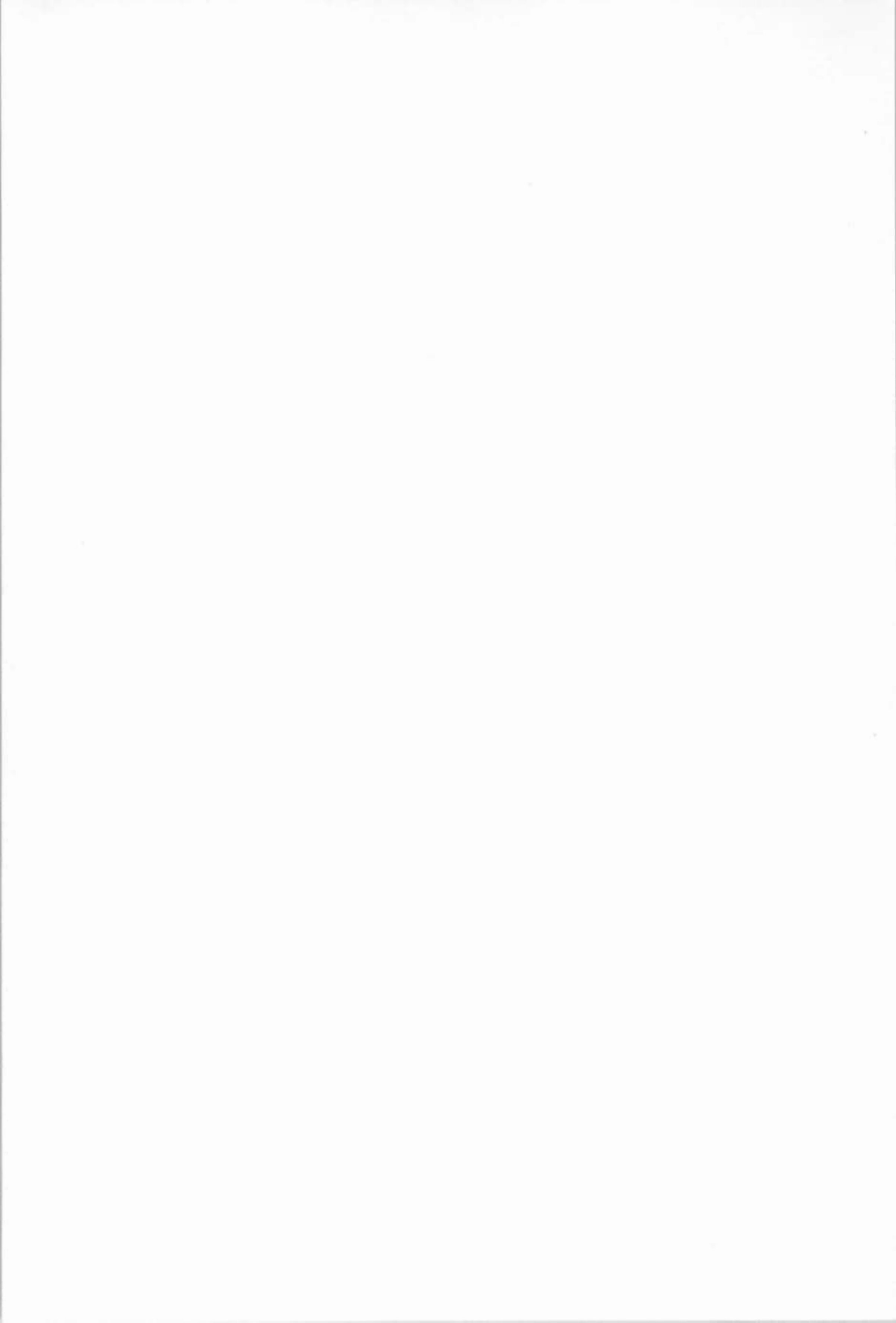
R_B Serienwiderstand im Gleichstromkreis einer Tunneldiode
R_D Betrag des negativen Widerstandes der Tunneldiode
R_G Generatorwiderstand
R_H HALL-Konstante
R_L Lastwiderstand
r Ortsvektor, Nummer der Nachbaratome, Rekombinationsrate
$r_{BB'}$ Basiswiderstand
r_C Ausgangswiderstand
r_E Eingangswiderstand
r_L Lastwiderstand
r' siehe Gl. 4.68
ϱ elektrische Ladungsdichte

s Oberflächenrekombinations-Geschwindigkeit
Σ_y' siehe Gl. 6.79
σ Leitfähigkeit
σ_N Leitfähigkeit im N-Gebiet
σ_n Elektronenanteil der Leitfähigkeit
σ_0 Gleichgewichtswert der Leitfähigkeit
σ^0 Leitfähigkeit an der Oberfläche
σ_{PE} Leitfähigkeit im Emitter
σ_p Löcheranteil der Leitfähigkeit
$\overline{\sigma_1}$ mittlere Leitfähigkeit der Basisschicht von $x=0$ bis $x=W_1$ (Abb. 6.25)
$\overline{\sigma_2}$ mittlere Leitfähigkeit der Basisschicht zwischen $x=W_1-W$ und $x=W_1$ (Abb. 6.25)

T absolute Temperatur
T_a Temperatur der Umgebung
T_i Temperatur am Kollektorübergang
t Zeit
τ_C Laufzeit der Löcher durch die Basis
τ_{eff} effektive Lebensdauer
τ_n Elektronenlebensdauer
τ_p Löcherlebensdauer
τ_s Stoßzeit
ϑ siehe Gl. 3.3

U	potentielle Energie, Spannung
U_{av}	Avalanche-Spannung
U_{At}	Potential des isolierten Atoms
U_B	Batteriespannung
U_{CB}	Kollektor-Basis-Spannung
U_{CB}'	Kollektor-Basis-Spannung am inneren Transistor
U_{CE}	Kollektor-Emitter-Spannung
U_{EB}	Emitter-Basis-Spannung
U_{EB}'	Emitter-Basis-Spannung am inneren Transistor
U_L	Leerlaufspannung
$U_{\max}$	maximal zulässige Spannung
U_0	Gleichspannungskomponente
U_{pt}	punch-through-Spannung
U_T	$\equiv kT/e$
$\overline{U}$	mittlerer Spannungsabfall an der Basis
$u(x, k)$	Amplitudenfaktor der Bloch-Funktion
u_{BE}	Basis-Emitter-Wechselspannung
u_{CB}	Kollektor-Basis-Wechselspannung
u_{CB}'	Kollektor-Basis-Wechselspannung am inneren Transistor
u_{CE}	Kollektor-Emitter-Wechselspannung
u_{EB}	Emitter-Basis-Wechselspannung
u_{EB}'	Emitter-Basis-Wechselspannung am inneren Transistor
u_1	Wechselspannungsamplitude
V	Volumen
V_D	Diffusionsspannung
V_L	Leistungsverstärkung, bezogen auf die maximal verfügbare Generatorleistung
$\overline{V_L}$	Leistungsverstärkung, bezogen auf die tatsächliche Eingangsleistung
V_{Lopt}	unilaterale, optimale Leistungsverstärkung
V_U	Spannungsverstärkung
V_I	Stromverstärkung
v	Driftgeschwindigkeit
v_g	Gesamtgeschwindigkeit
v_{lim}	Grenzgeschwindigkeit
v_n	Driftgeschwindigkeit der Elektronen
v_p	Driftgeschwindigkeit der Löcher
v_{th}	thermische Geschwindigkeit
W	Basisdicke
W_0	Basisdicke bei $U_{CB}' = 0$
W_1	siehe Abb. 6.10b, siehe Abb. 6.25
W_2	siehe Abb. 6.10b
w	Tunnelabstand
x	Ortskoordinate
x_N	Ende der Raumladungszone auf der N-Seite
x_P	Ende der Raumladungszone auf der P-Seite
x_L	siehe Abb. 5.41
$\overline{x}$	Ort des Wellenpakets
(Y')	Leitwertsmatrix des inneren Transistors
y	Ortskoordinate

y_{11} y_{11B} y_{12} y_{12B} y_{21} y_{21B} y_{22} y_{22B}	Vierpol-Elemente der Leitwertsmatrix in Basis-Schaltung
y'_{11} y'_{12} y'_{21} y'_{22}	Vierpol-Elemente der Leitwertsmatrix des inneren Transistors in Basisschaltung
y_{11E} y_{12E} y_{21E} y_{22E}	Vierpol-Elemente der Leitwertsmatrix in Emitterschaltung
y_E	Eingangsleitwert
y_G	Generatorleitwert
y_L	Lastadmittanz
z	Ortskoordinate
Z	Tunnelwahrscheinlichkeit



Erstes Kapitel

Charakteristische Eigenschaften und Definition eines Halbleiters

Bei der Ordnung der festen Stoffe nach ihrer elektrischen Leitfähigkeit stellt man fest, daß diese bei den verschiedenen Substanzen bei Zimmertemperatur mehr als 20 Größenordnungen umfaßt. Eine verhältnismäßig große Stoffgruppe, die der Metalle, steht am Anfang dieser Skala, beginnend mit einer spezifischen Leitfähigkeit von einigen 10^5 $Ohm^{-1}cm^{-1}$. Gute Isolatoren haben eine Leitfähigkeit von 10^{-17} $Ohm^{-1}cm^{-1}$ und weniger. Die Stoffe, die zu den Halbleitern gezählt werden, haben Leitfähigkeiten, die im allgemeinen zwischen denen der Metalle und denen der Isolatoren liegen.

Für eine Charakterisierung dieser Stoffgruppe der Halbleiter ist die Größe der elektrischen Leitfähigkeit allein jedoch nicht ausreichend. Weder die Elektrolyte noch Stoffe mit unzureichend definierter Zusammensetzung, wie nasser Bindfaden oder Beton (im LANDOLT-BÖRNSTEIN von 1927 noch unter Halbleitern aufgeführt), gehören zu den Halbleitern, wenn auch ihre Leitfähigkeit in der Größe der Halbleiterleitfähigkeit liegt. Träger der elektrischen Leitfähigkeit in Festkörpern sind allgemein Elektronen oder Ionen, manchmal beides. Ionenleiter wie die Alkalihalogenide bei höherer Temperatur haben mit den Halbleitern zwar viele gemeinsame Züge, trotzdem zählt man sie heute nicht mehr zu den Halbleitern. Unter Halbleitern versteht man demnach nur noch *elektronische* Halbleiter, d. h. solche halbleitende Festkörper, deren elektrische Leitfähigkeit durch Elektronen zustande kommt. Dabei ist es nicht immer leicht festzustellen, ob ein Ionenleiter oder ein Elektronenleiter vorliegt. Die Ionenleitung offenbart sich jedoch im allgemeinen durch einen deutlich nachweisbaren Materietransport beim Stromdurchgang.

Die Leitfähigkeit der Metalle nimmt mit wachsender Temperatur ab; bei Halbleitern ist es *meist* umgekehrt. Ein positiver Temperaturkoeffizient der elektrischen Leitfähigkeit genügt zur Definition eines Halbleiters jedoch auch nicht, weil viele Halbleiter Temperaturbereiche aufweisen, in denen die elektrische Leitfähigkeit mit der Temperatur abnimmt.

Zu einer zweckmäßigen Definition der Halbleiter gelangt man, wenn man die physikalischen Ursachen ihrer elektrischen Leitfähigkeit aufzeigt. Bei

Elektronenleitung ist die Leitfähigkeit bestimmt durch die Dichte der Elektronen (oder Leitfähigkeitsträger) und ihre Beweglichkeit, d. h. die mittlere Geschwindigkeit in Feldrichtung, die sie in einem Feld von 1 Volt/cm annehmen. Die hohe Leitfähigkeit der Metalle hat ihre Ursache darin, daß pro Atom etwa ein Leitfähigkeitselektron zur Verfügung steht. Diese Zahl der Elektronen in Metallen ist unveränderlich, eine Erwärmung oder Bestrahlung mit Licht ändert sie nicht. Ein Einbau geringer Mengen von Fremdstoffen beeinflußt vielleicht etwas die Beweglichkeit der Leitfähigkeitselektronen, jedoch nicht ihre Zahl. Auch die relativ geringe Abnahme der Leitfähigkeit der Metalle mit steigender Temperatur ist nur eine Folge der größeren Behinderung der Elektronen durch die größer werdenden Amplituden der Gitteratome, d. h. einer Abnahme der Beweglichkeit und nicht einer Änderung der Zahl der Elektronen.

Demgegenüber ist ein charakteristisches Merkmal der Halbleiter die außerordentliche Empfindlichkeit vieler Eigenschaften gegen geringe Zusätze von Fremdstoffen. So kam es in den vergangenen Jahren oft vor, daß die von zwei Forschern unabhängig hergestellten Proben desselben Materials beispielsweise in ihrer elektrischen Leitfähigkeit um mehrere Größenordnungen verschieden waren. Es ist völlig unmöglich, solche Erscheinungen nur mit verschiedenen Beweglichkeiten der Elektronen zu erklären. Ebensowenig ist die Leitfähigkeitszunahme der Halbleiter um viele Größenordnungen in manchen Temperaturbereichen nur als Änderung der Beweglichkeit zu verstehen. Man ist deshalb gezwungen — und die diesbezüglichen Experimente bestätigen dies —, solche Änderungen einer *veränderlichen Dichte der Leitfähigkeitsträger* zuzuschreiben. Selbstverständlich kann auch die Beweglichkeit der Elektronen im Halbleiter geändert werden. Dies tritt aber oft gegen die Trägerdichteänderung völlig zurück. Bekannt ist die Änderung der elektrischen Leitfähigkeit vieler Halbleiter durch Bestrahlung mit Licht oder Korpuskular-Strahlen, die auch auf einer Änderung der Elektronendichte beruht.

Diese Sachverhalte ermöglichen nun eine einwandfreie Definition des (elektronischen) Halbleiters: Er ist ein Stoff, dessen Leitfähigkeitsträgerdichte durch äußere Einflüsse, wie z. B. Temperatur, elektromagnetische oder Korpuskular-Strahlen, thermische Behandlung und Zusätze geringer Mengen anderer Stoffe, erheblich verändert werden kann.

Zweites Kapitel

Germanium als einfachster Halbleitertyp

2.1 Vorbemerkung

Für eine Einführung in den elektrischen Leitungsmechanismus eines Halbleiters ist die Betrachtung des Elements Germanium besonders gut geeignet. Hier findet man die klarsten Verhältnisse vor; außerdem gehört dieser Stoff heute zu den am besten untersuchten Halbleitern. Wenn trotz über tausend zum Teil grundlegender Arbeiten auch bei diesem Stoff bisher viele Fragen nur teilweise beantwortet werden können, so zeigt das, wie kompliziert im Grunde ein Festkörper ist.

Wenn wir hier dem Germanium als „Schulstoff" den Vorrang vor den Alkalihalogeniden geben, die dank der Arbeiten von R. W. Pohl [1] und seinen Mitarbeitern eigentlich am Anfang der quantitativen Halbleiterphysik stehen, so findet das seine Rechtfertigung darin, daß bei uns in erster Linie Phänomene betrachtet werden, die mit dem elektrischen Leitungsmechanismus zusammenhängen. Auch für den PN-Schichtkristall liegen beim Germanium besonders einfache Verhältnisse vor, so daß dieser Stoff gleichzeitig auch ein Schulbeispiel für die Erläuterung der Wirkungsweise der Gleichrichter und Transistoren ist.

Der Leitungsmechanismus des Germaniums ist grundsätzlich bei vielen Halbleitern realisiert. Dem Germanium am ähnlichsten ist das Silizium. Es ist von immer noch wachsender Bedeutung als technischer Halbleiter. Aus diesen Gründen werden wir im folgenden verschiedentlich auch Bemerkungen über das Silizium zu machen haben. Daneben zeigen viele intermetallische Verbindungen zwischen Elementen der III. und V. Gruppe, wie z. B. Galliumarsenid und Indiumantimonid, die meisten Alkalihalogenide, einige Sulfide und Oxyde, ähnliche Verhältnisse. Ob aber alle Halbleiter, z. B. Tellur, Selen und Bor, zwanglos in diesem einfachen Schema unterzubringen sind, ist eine noch offene Frage.

Nicht unwesentlich für die Beantwortung der Frage, warum gerade Germanium so häufig untersucht wird, ist der Gesichtspunkt, daß dieser Stoff verhältnismäßig leicht gereinigt und als Einkristall dargestellt werden kann. Ein Schulstoff muß nicht nur „einfach" in seinem Verhalten, sondern auch experimentell realisierbar sein!

2.2 Zur Technologie des reinen Germaniumeinkristalls

2.2.1 Reinigung. Bei der chemischen Analyse eines silberreichen Erzes aus einer Grube bei Freiberg/Sachsen stellte CL. WINKLER immer einen Fehlbetrag von 6 bis 7% fest. Er ist dieser Sache nachgegangen und konnte im Jahre 1886 ein neues Element, das Germanium, isolieren, das 15 Jahre früher bei der berühmten Ordnung der Elemente von MENDELEJEFF noch vermißt und mit dem Namen Ekasilicium versehen worden war. Bis heute ist das Germanium relativ selten geblieben. Es kommt im wesentlichen als Beimengung in sulfidischen Erzen im Kongo, in Südwestafrika, den USA und der Sowjetunion vor.

Zur Reindarstellung strebt man mit Hilfe chemischer Methoden meist eine flüssige Germaniumverbindung an, beispielsweise $GeCl_4$, das dann durch Destillation und Rektifikation nachgereinigt wird. Das Chlorid wird mit reinem Wasser hydrolysiert und das als weißes Pulver gewonnene Germaniumoxyd mit reinem Wasserstoff bei etwa 850°C reduziert. Das so dargestellte Germanium fällt als ein mausgraues, mehliges Pulver an. Dieses kann (u. U. im selben Ofen) durch Erhitzen auf 1000°C (Schmelzpunkt von Ge: 937°C) in Graphitformen zu Barren geschmolzen werden. Unter Anwendung der bei chemischen Reindarstellungen üblichen Sorgfalt erhält man ein spektralreines Material, d. h. man hat die Fremdstoffe auf weniger als 10^{-5} bis 10^{-6} Atomteile reduziert. Auf 10^5 bis 10^6 Germaniumatome kommt demnach ein Fremdstoffatom.

Als Ausgangsmaterial für Dioden und Transistoren liegt aber heutzutage in jeder Halbleiterfabrik Germanium auf Lager, das um etwa fünf Größenordnungen reiner ist. Das erreicht man mit einer physikalischen Methode, die auf der verschiedenen Löslichkeit der Fremdstoffe im festen und flüssigen Germanium beruht. Aus dem Zustandsdiagramm eines Zweistoffsystems ist zu entnehmen, daß im allgemeinen der erstarrte Teil einer Schmelze eine andere Zusammensetzung besitzt als die flüssige Phase. Dies gilt auch noch, wie nachgewiesen wurde, für geringste Konzentrationen von Fremdstoffen in Germaniumschmelzen. Bis auf wenige Ausnahmen ist es hier so, daß die Löslichkeit der Fremdstoffe in der festen Phase viel geringer ist als in der flüssigen. Wenn in einem länglichen Tiegel oder „Boot" flüssiges Germanium von einem Ende her erstarrt (Abb. 2.1), so enthält demnach das erstarrte Material den Fremdstoff immer in geringerer Konzentration als die Schmelze. Bei fortschreitender Erstarrung der Schmelze nimmt deshalb der Fremdstoffgehalt des festen Kristalls in Abb. 2.1 von links nach rechts zu. Wenn man das rechte Kristallende, das reich an Störstoffen ist, entfernt, den Kristall neu aufschmilzt und wieder von einem Ende her erstarren läßt, erhält man einen reineren Kristall. Dieses alte Verfahren ist jedoch sehr umständlich.

Einfacher ist es, wenn durch einen festen Kristall eine flüssige Zone bestimmter Länge immer in derselben Richtung hindurchgeschoben wird [2]. Dabei

ist der wiedererstarrte linke Teil immer ärmer an Fremdstoffen als der noch nicht aufgeschmolzene rechte Teil des Kristalls. Eine beliebig häufige Wiederholung dieses Reinigungsschrittes und gleichzeitiges Durchschieben mehrerer flüssiger Zonen gemäß Abb. 2.2 ist möglich, ohne daß der Kristall aus dem Tiegel genommen werden muß, und führt bei Germanium schließlich im größten Teil des Kristalls zu der bereits genannten Reinheit von weniger als 10^{-10} Atomteilen Fremdstoffgehalt. Eine technische Ausführung dieses Verfahrens ist in Tafel 1 gezeigt.

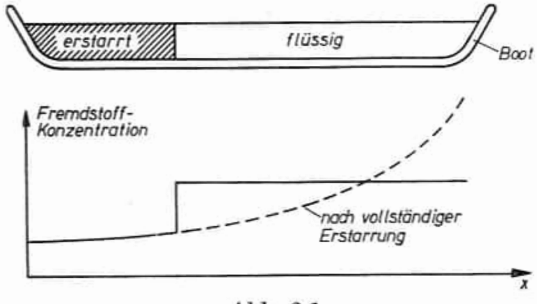


Abb. 2.1.
Erstarrung einer Germaniumschmelze und schematischer Verlauf der Fremdstoffkonzentration.

Obwohl die Methode, Stoffe durch Umschmelzen oder Umkristallisieren zu reinigen, nicht neu ist, hat das oben beschriebene Zonenreinigungsverfahren die

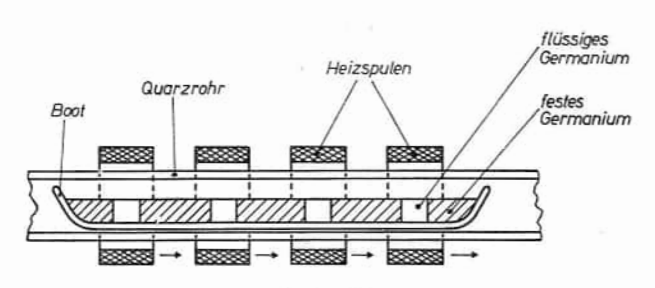
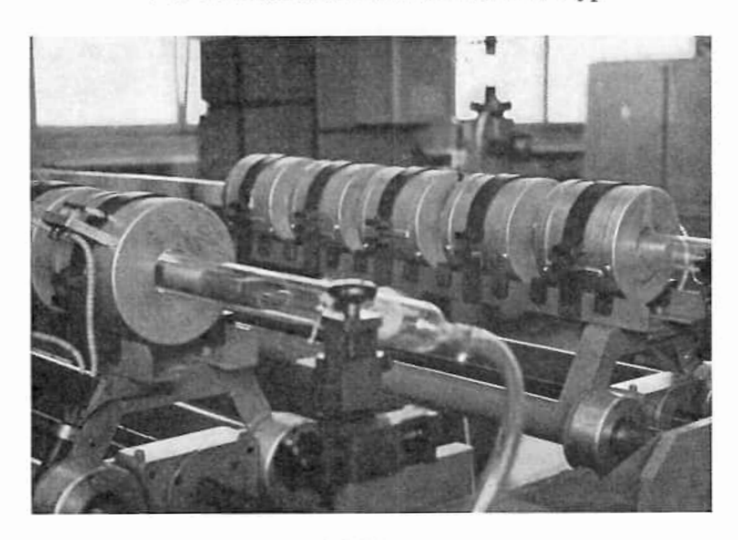


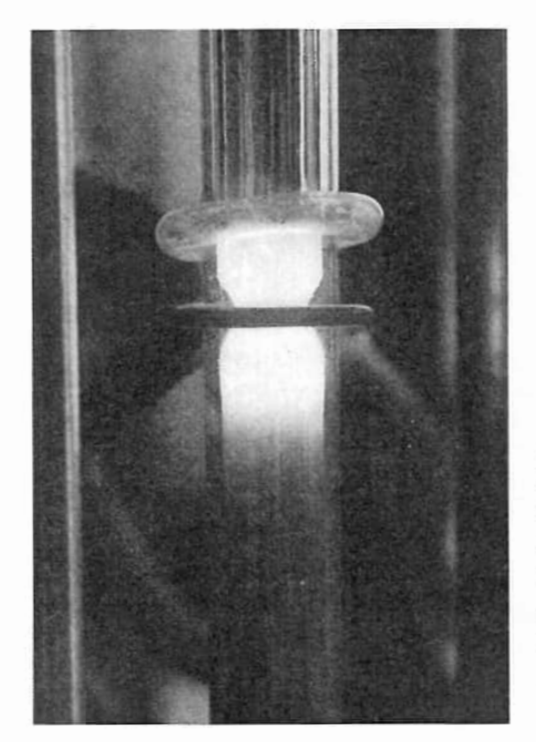
Abb. 2.2.
Zonenreinigung von Germanium.

Reindarstellung der Halbleiter entscheidend gefördert. Das Verfahren wird neuerdings auch bei Metallen, anorganischen und insbesondere organischen Stoffen erfolgreich angewendet. Voraussetzung dafür ist, daß die Störstoffe in den verschiedenen Phasen verschieden löslich sind und der Tiegel keine Störstoffe abgibt oder mit dem Halbleiter reagiert. Für die Zonenreinigung von Germanium eignen sich am besten Boote aus reinstem Graphit oder kohleüberzogenem Quarz.

Die Reaktion des Halbleiters mit dem Tiegel läßt sich ganz vermeiden, wenn man einen Stab vertikal an beiden Enden fest einspannt, in einer inerten Atmosphäre oder im Vakuum mit Hilfe der Induktionsspule eines HF-Generators eine kurze Zone aufschmilzt und wie bei der Zonenreinigung durch den Stab führt [3, 4, 5]. Falls der Stab nicht zu dick ist, verhindert die Ober-



Tafel 1.
Technische Ausführung einer Zonenreinigungsanlage für Germanium (Werkphoto Intermetall).



Tafel 2.
Detailaufnahme einer tiegelfreien Zonenreinigungsanlage in Betrieb. Der das senkrechte Quarzrohr umgebende obere Ring ist eine Luftdusche, die zur Kühlung der Quarzwand dient. Sie besteht aus einem durchlöcherten Glasrohr und wird mit Preßluft gespeist. Der untere Ring ist die aus einer Windung bestehende Induktionsspule, ein wassergekühltes Kupferrohr.

flächenspannung des aufgeschmolzenen Teils, zusammen mit den elektromagnetischen Kräften des Hochfrequenzfeldes, das Herunterfließen. Wichtiger als für Ge ist diese Methode für die Zonenreinigung von Silizium, da hier wegen der höheren Schmelztemperatur (1420°C) die Reaktion mit dem Tiegel besonders störend ist. Abb. 2.3 und Tafel 2 zeigen eine Anordnung zum tiegelfreien Zonenschmelzen von Silizium. Bei Zonenlängen von etwa 1 cm kann man noch Stäbe von 1 cm Durchmesser und mehr reinigen. Das Kardinalproblem bei der Reinigung von Silizium, nämlich die Entfernung des Bors, ist wegen seiner nahezu gleichen Löslichkeit in der flüssigen und festen Phase mit dieser Methode allerdings nicht zu lösen.

2.2.2 Einkristallzucht. Für die Herstellung von Halbleiterbauelementen, wie hochsperrende Gleichrichter und Transistoren, wird einkristallines Material benötigt, und zwar aus folgenden Gründen: In den aktiven Bereichen dieser Bauelemente fallen über kürzeste Entfernungen hohe Spannungen ab. Um dabei unterhalb der kritischen Durchschlagsfeldstärke zu bleiben, ist höchste Gleichmäßigkeit des Grundmaterials nötig. Korngrenzen sind aber nicht nur eine Unterbrechung in der regelmäßigen Gitterstruktur des Materials, sondern sie bilden auch Anreicherungszonen für Verunreinigungen. Außerdem wird der für die Transistorwirkung entscheidende Fluß der Stromträger vom Emitter zum Kollektor durch Korngrenzen im Kristall empfindlich gestört.

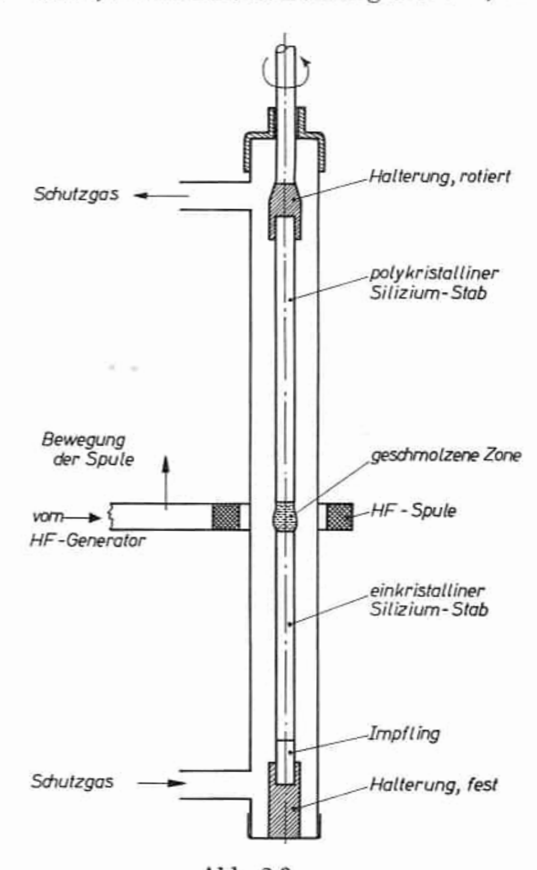


Abb. 2.3.
Tiegelfreies Zonenschmelzen (Floating Zone) von Silizium.

Schließlich leuchtet ein, daß ein Einkristall wegen seiner größeren Einfachheit sich viel eher zu grundsätzlichen Untersuchungen eignet als ein Vielkristall. Wenn man z. B. heute, nachdem der Selengleichrichter bereits 40 Jahre im Gebrauch ist, noch heftige Diskussionen über den Leitfähigkeits-

und Gleichrichtermechanismus des Selens führt, so liegt das mit daran, daß es noch nicht gelungen ist, einkristallines Selen mit einer ähnlichen Fehlordnung, wie es im Selengleichrichter vorliegt, herzustellen.

Es ist hier nicht der Ort, alle verschiedenen Methoden zur Herstellung von Einkristallen zu erörtern; wir wollen uns damit begnügen, auf ein besonders erfolgreiches Verfahren hinzuweisen. Abb. 2.4 zeigt eine moderne Durchführung der CZOCHRALSKI-Methode [6, 7, 8, 9]: Ein Tiegel, der von einem Quarzrohr umgeben ist, enthält das Schmelzgut, z. B. Germanium, das von einem hochfrequenten Wechselfeld geschmolzen wird. Das Wechselfeld wird von einer Spule erzeugt, die am Ausgang eines kräftigen HF-Generators liegt. Bei einer Temperatur knapp oberhalb der Schmelztemperatur wird von oben ein Impfling (d. h. ein einkristallines Stück Ge) in die Schmelze getaucht und langsam nach oben gezogen, wobei sich festes Germanium einkristallin an den Keim anlagert. Ein gleichmäßiges Anwachsen des Schmelzgutes erreicht man am einfachsten durch Drehung des Kristalls. Eine zusätzliche Drehung des Tiegels beseitigt Unsymmetrien in der Temperaturverteilung des Tiegels. Heißes Germanium und Silizium sind in festem Zustand brillanter, erscheinen also heller als die Schmelze, ein Umstand, der das Züchten dieser Kristalle erleichtert. Man kann deshalb bei visueller Beobachtung die Dicke des wachsenden Kristalles regulieren, indem man mehr oder weniger Heizenergie zuführt. Auch mit der tiegelfreien Methode (Abb. 2.3) lassen sich Einkristalle gleicher Dicke herstellen, wenn man an ein Ende einen einkristallinen Impfling setzt.

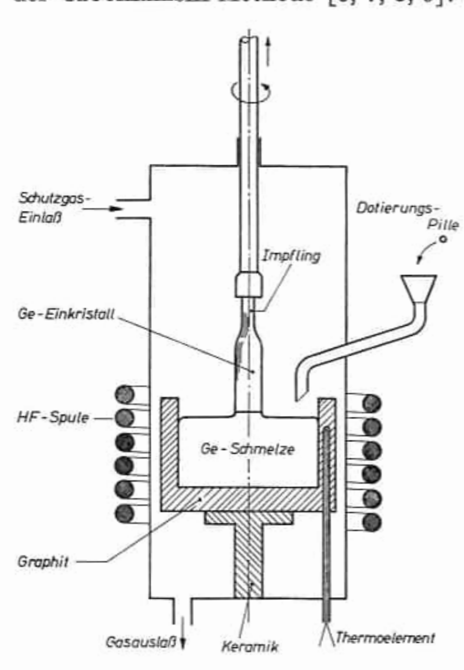
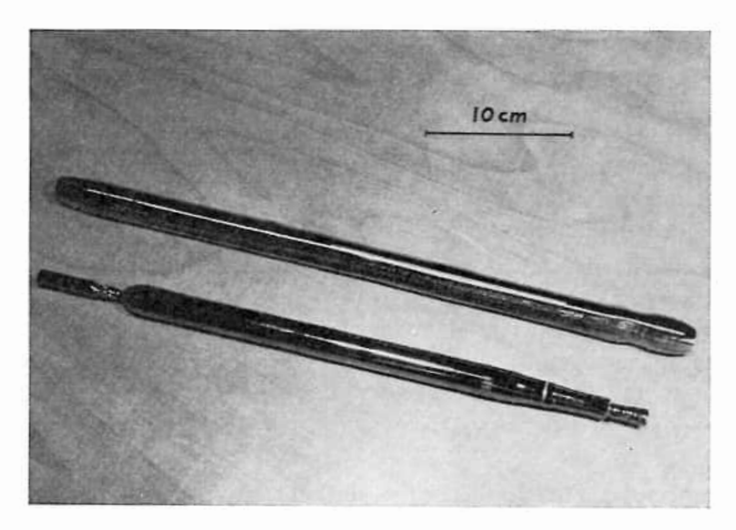


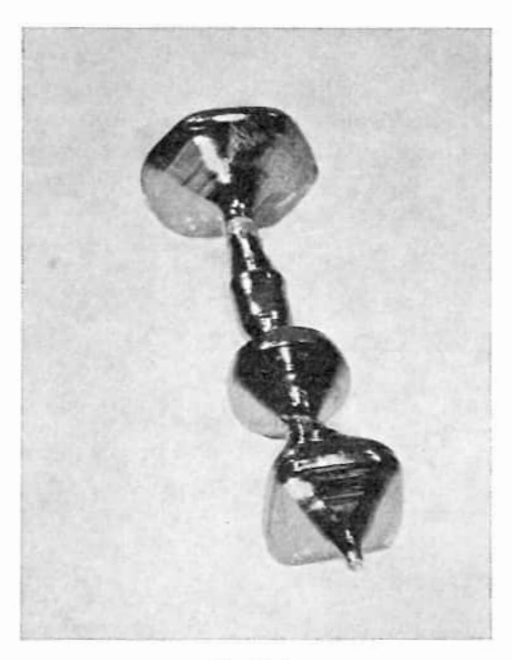
Abb. 2.4.
Einkristallzucht von Germanium nach CZOCHRALSKI.

Die in Tafel 3 gezeigten Kristalle (Germanium nach CZOCHRALSKI — oben, Silizium tiegelfrei — unten) sollen illustrieren, daß man mit dieser Methode Einkristalle gleichmäßiger Dicke herstellen kann. Um die Kristallsymmetrie gut hervortreten zu lassen, wurde bei einem Kristall (Tafel 4) die Temperatur der Schmelze so variiert, daß der Kristall seine Dicke änderte. Germanium kristallisiert im kubischen System des Diamantgitters. Die [100]-Richtung war hier zunächst Zugrichtung des Kristalls (dem Betrachter zugekehrtes



Tafel 3.

Ein Germaniumkristall, nach CZOCHRALSKI [6] aus dem Tiegel gezogen (oben), und ein Siliziumkristall, tiegelfrei gezogen (unten). Die dünneren Teile an den Enden des Siliziumstabes sind der Impfling bzw. das obere Einspannstück (Werkphoto Intermetall).



Tafel 4.

Symmetrie eines Germaniumkristalls (Werkphoto Intermetall).

Ende). Man sieht deutlich die Vierzähligkeit dieser Achse. Die {111}-Ebenen sind beim Diamantgitter wachstumsgehemmt; das zunächst entstandene oktaederähnliche Gebilde ist durch Schnitte dieser {111}-Flächen mit der Oberfläche der Schmelze entstanden. Bei weiterem Ziehen wurde der Stab polykristallin (runder Querschnitt), und erst zum Schluß gelang wieder einkristallines Wachstum, das jetzt aber in <111>-Richtung erfolgte. Dadurch entstand ein etwa sechseckiger Querschnitt.

Interessant ist, daß nicht nur beim Aufbau des Germaniumkristalls die {111}-Ebene wachstumsgehemmt ist. Auch der Abbau des Einkristalls durch Auflösen, z. B. mit Hilfe von Metallen wie Indium oder Aluminium, wird in dieser Richtung gehemmt, ein Umstand, den man bei der Herstellung von Gleichrichtern und Transistoren nach dem Legierungsverfahren ausnützt (siehe Abschnitt 5.3.2 und Tafel 15).

2.2.3 Störungen im Einkristall. Selbst die mit großer Sorgfalt gezüchteten reinen Einkristalle sind nicht vollständig störungsfrei. In jedem Kristall treten Gitterbaufehler auf, von denen die folgenden drei Arten erwähnt seien:

a) Normale Gitterplätze sind unbesetzt (SCHOTTKYsche Fehlordnung), oder Gitteratome haben ihre normalen Gitterplätze verlassen und besetzen Plätze zwischen den Gitteratomen (FRENKELsche Fehlordnung).

b) Fremdatome sind auf normalen Gitterplätzen oder auf Zwischengitterplätzen eingebaut.

c) Einzelne Netzebenen des Einkristalls sind gegeneinander verschoben. Zu dieser Art von Gitterbaufehlern gehören die Versetzungen. In Abb. 2.5 ist eine Stufenversetzung schematisch dargestellt, bei der durch einen äußeren

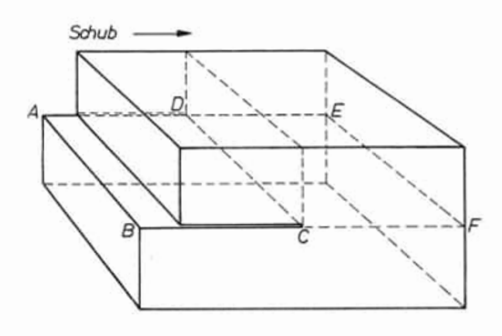


Abb. 2.5.

Stufenversetzung bei CD, entstanden durch Stauchen des oberen Kristallteils gegenüber dem unteren um einen Atomabstand. ABEF ist Gleitebene.

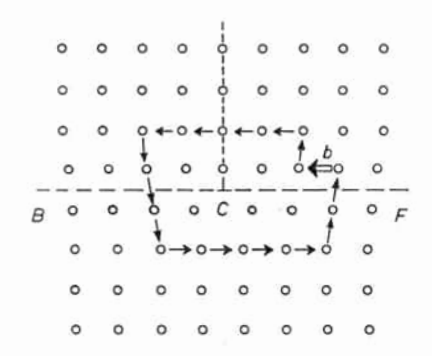


Abb. 2.6.

Stufenversetzung bei C, entstanden durch Einschieben einer Halbebene von oben.

mechanischen Einfluß die obere Kristallhälfte parallel zur unteren um einen Atomabstand gestaucht worden ist. Entsprechend Abb. 2.6 läßt sich die Stufenversetzung auch mit Hilfe einer eingeschobenen Halbebene beschreiben. Bei einer Schraubenversetzung ist ein Teil des oberen Kristallteils gemäß Abb. 2.7 gegen den unteren Teil verschoben.

Die genannten Gitterbaufehler entstehen bei der Herstellung der Einkristalle. SCHOTTKYsche bzw. FRENKELsche Fehlordnung ist in jedem Kristall wegen der statistischen Wärmebewegung der Atome am Erstarrungspunkt in großer Zahl vorhanden. Beim Abkühlen auf Zimmertemperatur wird ein mehr oder weniger großer Teil davon eingefroren. Fremdatome können, soweit sie nicht schon in der Schmelze sind, auch durch das verwendete Tiegelmaterial in den Kristall eindringen und Gitteratome von ihren Plätzen verdrängen. Versetzungen entstehen u. a. infolge ungleichmäßiger Abkühlung des Kristalls, da die dabei auftretenden thermischen Spannungen zu einer plastischen Verformung des Kristalls führen [10]. Schließlich können während des Ziehvorgangs Versetzungen aus dem Impfling in den Einkristall wandern.

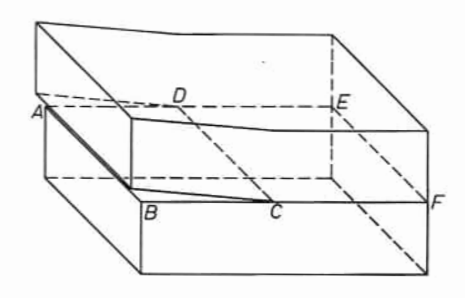


Abb. 2.7. Schraubenversetzung

Praktisch wird man die Ausbildung von Gitterbaufehlern während des Ziehprozesses nie vollständig verhindern, jedoch kann man durch besondere Maßnahmen den gezogenen Einkristall möglichst störungsarm machen. So ist es in neuester Zeit gelungen, sogar versetzungsfreie Kristalle herzustellen [11].

Die Auswirkungen der Gitterbaufehler auf das elektrische Geschehen in einem Halbleiter können an dieser Stelle nur angedeutet werden. Die Gitterbaufehler (insbesondere Typ b) verändern die Trägerdichte und in geringerem Maße die Beweglichkeit. Die Versetzungen wirken, ähnlich wie die oben erwähnten Korngrenzen, insbesondere als Rekombinationszentren auf die sogenannten Minoritätsladungsträger in einem Halbleiter, die von großer Bedeutung für das Transistorgeschehen sind. Bei der Herstellung von Bauelementen durch Legieren oder durch Diffusion (siehe Abschnitt 5.3) wirkt sich die Versetzungsdichte auf die Gleichmäßigkeit und Beherrschbarkeit dieser Prozesse aus.

Die Güte eines Halbleitereinkristalls ist daher unter anderem auch durch die Anzahl der Versetzungen charakterisiert. Die Bestimmung der Versetzungsdichte erfolgt durch Auszählen, nachdem man die Versetzungen sichtbar gemacht hat mit Hilfe einer geeigneten Ätzlösung [10, 12], die die gestörten Kristallbereiche stärker abträgt als die ungestörten. An den Durchstoßpunkten

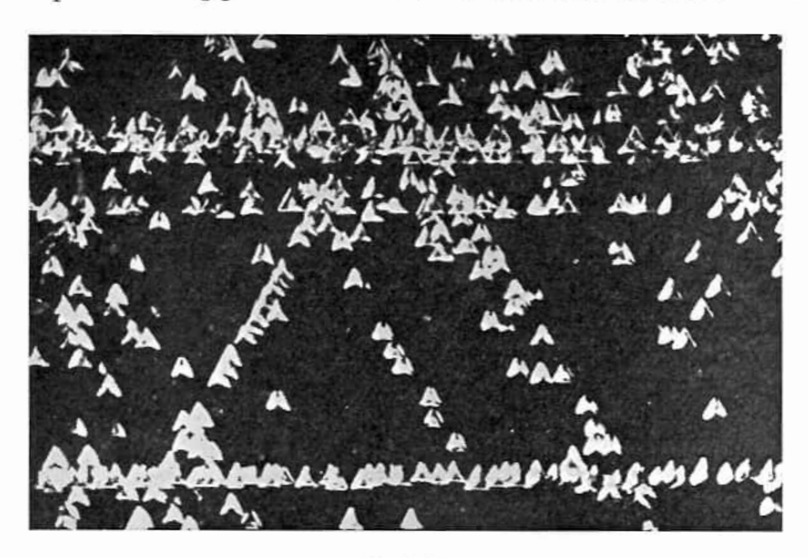
der Versetzungslinien durch die Oberfläche entstehen in einer solchen Lösung kleine Ätzgrübchen, die sogenannten „etch pits“ (Tafel 5).

Eingehendere Informationen über die Versetzungen in einem Einkristall bringt die Dekorierungsmethode [12, 13], mit deren Hilfe der Verlauf von Versetzungslinien im Kristallinneren sichtbar gemacht werden kann. Dieses Verfahren eignet sich besonders gut für Silizium, in das man bei hoher Temperatur Kupfer eindiffundiert. Die bei einer bestimmten Temperatur lösliche Kupfermenge sinkt mit fallender Temperatur; Cu muß daher während der Abkühlung ausgeschieden werden. Aus energetischen Gründen geschieht die Ausscheidung bevorzugt an Versetzungslinien, die dadurch dekoriert sind. Bei einer Ultrarotdurchleuchtung des Kristalls können die dekorierten Versetzungen mit Hilfe spezieller Filme oder eines Bildwandlers sichtbar gemacht werden (siehe Tafel 6).

2.3 Temperaturabhängigkeit von elektrischem Widerstand und HALL-Effekt

2.3.1 Leitfähigkeit. Mit einer Schaltung, wie sie in Abb. 2.8 schematisch skizziert ist, kann die elektrische Leitfähigkeit reiner und mit verschiedenen Zusätzen versehener Ge-Proben gemessen werden. Abb. 2.9 stellt die technische Form einer Anordnung dar, in der die Leitfähigkeit (und der HALL-Effekt — siehe unten) bei verschiedenen, insbesondere sehr tiefen Temperaturen gemessen wurden. Der Verlauf der Leitfähigkeit σ mit der Temperatur T ist in Abb. 2.10 in schematischer Weise angegeben. Tatsächliche Meßergebnisse sind in Abb. 2.11 und Abb. 2.12 dargestellt. Dabei wurde nicht die Leitfähigkeit, sondern deren Reziprokwert, der spezifische Widerstand, logarithmisch gegen die reziproke Temperatur aufgetragen. Abb. 2.11 gibt Messungen bis herab zu $11\,°K$ an Germanium wieder, das mit verschiedenen Mengen von Arsen dotiert worden ist. Abb. 2.12 zeigt Messungen an Gallium-dotiertem Germanium zwischen Zimmertemperatur und 900°C.

Man erkennt in diesen Kurven (2.10–2.12) drei Temperaturbereiche mit verschiedenem Verhalten. Bei tiefen Temperaturen steigt die Leitfähigkeit etwa exponentiell mit der Temperatur. Die Zusätze erhöhen hier allgemein die Leitfähigkeit mit der Wurzel der Störstoffkonzentration. Bei mittleren Temperaturen nimmt die Leitfähigkeit etwas mit der Temperatur ab. Größenordnungsmäßig ist die Leitfähigkeit hier proportional mit der Störstoffkonzentration. Da bei den Metallen ein ähnliches Verhalten vorliegt, charakterisieren wir diesen Bereich als metallischen Bereich. Bei höherer Temperatur ist die Leitfähigkeit unabhängig vom Störstoffgehalt und nimmt exponentiell mit der Temperatur zu. Hier offenbart sich die Leitfähigkeit des Halbleitermaterials selbst, wir nennen deshalb diesen Bereich Eigenleitfähigkeit. Bei tiefer Temperatur bestimmen die Störstoffe die Leitfähigkeit des Halbleiters, weshalb dieser Bereich Störleitungsbereich genannt wird.



Tafel 5.

Ätzgrübchen an einer (111)-Oberfläche eines Siliziumkristalls. Die Aufreihung der Versetzungen entlang den Seiten von gleichwinkligen Dreiecken ist auf Gleitungen in den (111)-Ebenen zurückzuführen.
(Vergrößerung etwa 70fach; nach RIEGER und SEILER [13]).



Tafel 6.

Dekorierte Versetzungen in Silizium.
(Vergrößerung etwa 100fach; nach RIEGER und SEILER [13]).

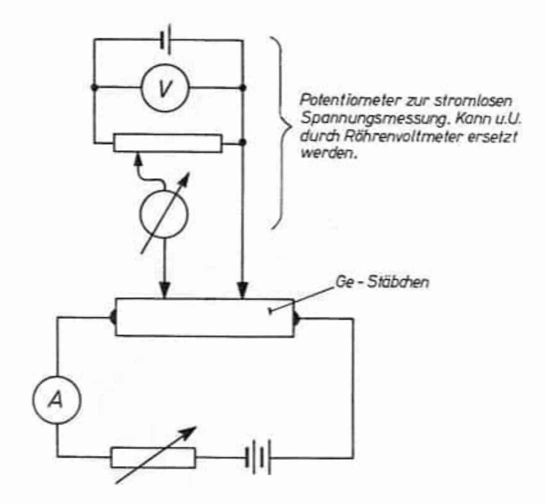


Abb. 2.8.
Schaltung für Leitfähigkeitsmessung.

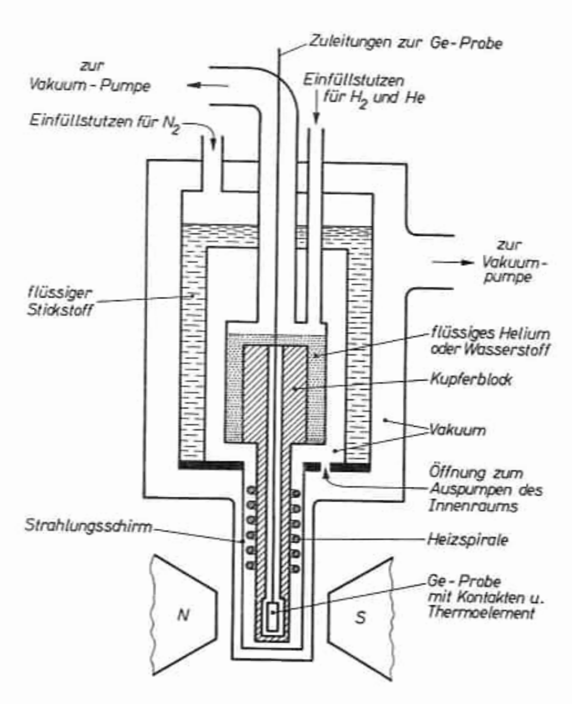


Abb. 2.9.
Anordnung zur Messung von Leitfähigkeit und Hall-Effekt zwischen 15°K und 400°K (nach W. C. Dunlap [14]). Die Apparatur ist ganz aus Metall.

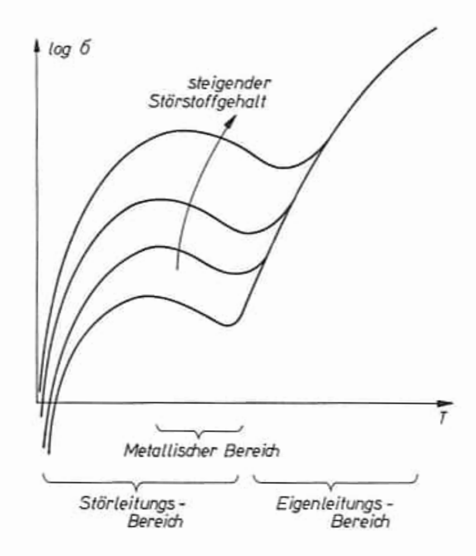


Abb. 2.10.
Schematischer Verlauf der Leitfähigkeit mit der Temperatur. Kurvenparameter sind die Verunreinigungskonzentrationen und zwar ansteigend in Pfeilrichtung.

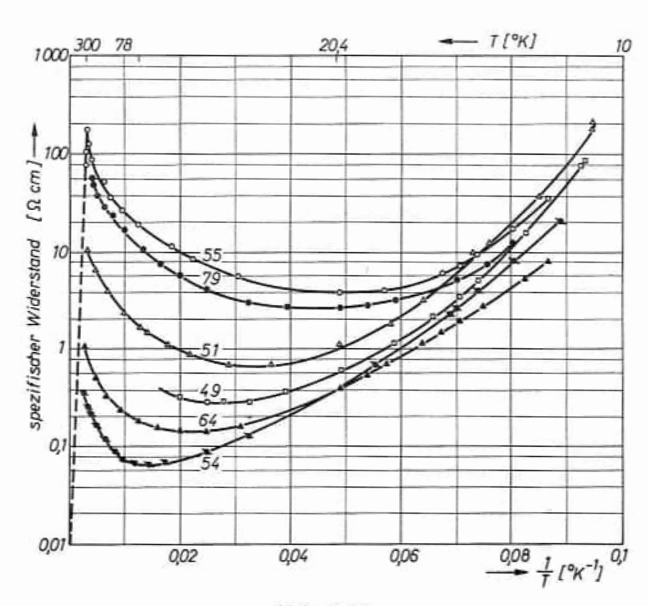


Abb. 2.11.
Spezifischer Widerstand als Funktion der Temperatur für verschieden stark mit Arsen dotiertes Germanium (nach DEBYE und CONWELL [15]).

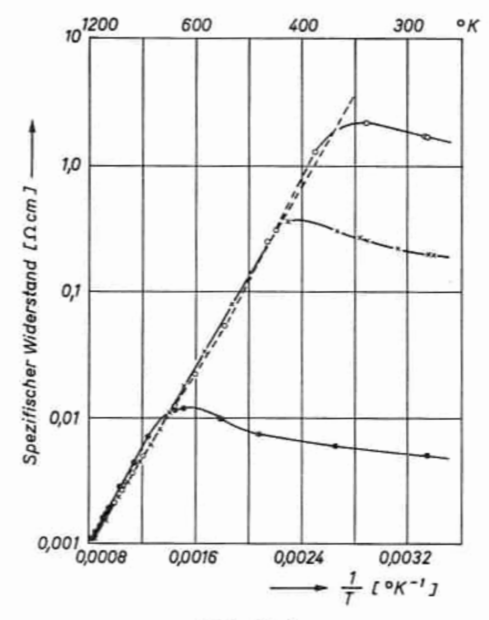


Abb. 2.12.
Spezifischer Widerstand als Funktion der Temperatur für verschieden stark mit Gallium dotiertes Germanium (nach DUNLAP, McMILLAN und BROOKS [16]).

Die Dichte des elektrischen Stromes $\boldsymbol{j}$, der von bewegten Leitfähigkeitsträgern — im folgenden oft Träger genannt — der Ladung q, Dichte p und Geschwindigkeit $\boldsymbol{v}$ hervorgerufen wird, ist

$$\boldsymbol{j} = q \cdot p \cdot \boldsymbol{v} \qquad \text{(Gl. 2.1)}$$

Nach dem Ohmschen Gesetz ist $\boldsymbol{v}$ der elektrischen Feldstärke $\boldsymbol{F}$ proportional

$$\boldsymbol{v} = \mu \cdot \boldsymbol{F} \qquad \text{(Gl. 2.2)}$$

μ nennt man die Beweglichkeit der Leitfähigkeitsträger. Die spezifische Leitfähigkeit σ — im folgenden auch oft kurz mit Leitfähigkeit bezeichnet — die $\boldsymbol{j}$ mit $\boldsymbol{F}$ verknüpft ($\boldsymbol{j} = \sigma \cdot \boldsymbol{F}$) ist also

$$\sigma = q \cdot p \cdot \mu \qquad \text{(Gl. 2.3)}$$

Wir wollen, wenn nichts anderes bemerkt wird, die Beweglichkeit als eine feld unabhängige Konstante betrachten (Gültigkeit des Ohmschen Gesetzes!). Dies ist für kleine Felder gerechtfertigt. Neuere Untersuchungen an Germanium und Silizium haben aber gezeigt, daß hier bei Feldern von einigen 1000 V/cm die Beweglichkeit der Träger feldabhängig wird.

Bei Betrachtung der Abb. 2.10–2.12 an Hand der Gl. 2.3 ergibt sich, daß die starken Änderungen der Leitfähigkeit kaum mit der Temperaturabhängigkeit der Beweglichkeit erklärt werden können. Über die möglichen Änderungen

dieser Größe ist man besonders von dem Studium der Metalle her orientiert. Es ist unmöglich, die Leitfähigkeitsänderungen im Eigenleitungsgebiet und bei tiefen Temperaturen des Störleitungsgebietes mit Beweglichkeitsänderungen zu deuten. Man wird so zu dem Schluß gezwungen, daß die Trägerdichte stark von der Temperatur abhängig ist.

Dieser Sachverhalt wird weiterhin untermauert und erklärt durch den nachfolgend beschriebenen HALL-Effekt.

2.3.2 HALL-Effekt. Der HALL-Effekt ist ein besonders wichtiges Hilfsmittel, um sowohl die Polarität der Leitfähigkeitsträger als auch ihre Dichte zu messen. Ein quaderförmiges Stück des zu untersuchenden Materials wird zu diesem Zweck an seinen Stirnseiten möglichst großflächig kontaktiert (Abb. 2.13). Diese Kontakte dienen als Stromzuführungen. Auf zwei anderen Flächen werden einander genau gegenüberliegend zwei kleine Potentialkontakte angebracht. Diese Anordnung wird so zwischen die Pole eines Magneten gebracht, daß das Magnetfeld senkrecht zum dritten nicht kontaktierten Flächenpaar steht (Abb. 2.14).

Die Träger des Stromes $\boldsymbol{I}$ (vgl. Abb. 2.13) werden in dem Magnetfeld, das senkrecht zur elektrischen Feldrichtung steht, gemäß der LORENTZ-Kraft

$$\boldsymbol{K} = q[\boldsymbol{v}\,\boldsymbol{B}] \qquad \text{(Gl. 2.4)}$$

solange nach der Seite abgedrängt, bis die durch die oberflächliche Ladungsansammlung entstehende Feldstärke die Ablenkung kompensiert. Man sieht sofort, daß in Abb. 2.13 für positive Stromträger (Ladung $+\ q > 0$), bei denen $\boldsymbol{v}$ die Richtung von $\boldsymbol{I}$ hat, eine Kraft $\boldsymbol{K}$ nach oben zustande kommt. Sie führt zu einer Anhäufung positiver Ladung an der oberen Fläche und damit zu einem Querfeld von oben nach unten. Für negative Stromträger ist in Gl. 2.4 q durch $-q$ und v durch $-v$ zu ersetzen. Dies ergibt ebenfalls eine Kraft $\boldsymbol{K}$ nach oben. Sie führt nun aber zu einer Anhäufung negativer Ladung an der oberen Fläche und damit zu

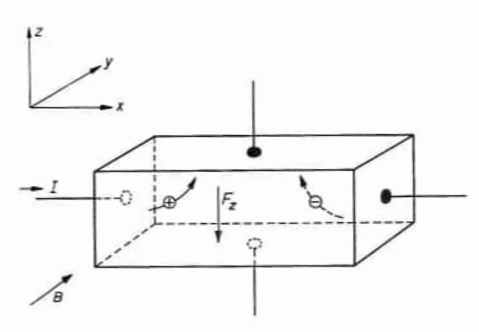


Abb. 2.13.
Zur Erklärung des HALL-Effekts.

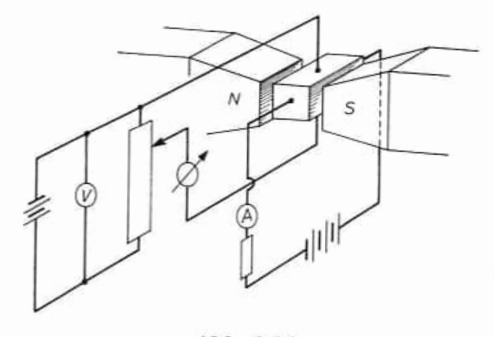


Abb. 2.14.
HALL-Effekt-Messung.

einem Querfeld von unten nach oben. Aus dem Vorzeichen der durch die magnetische Ablenkung entstehenden Spannung an den Potentialsonden, der HALL-Spannung, läßt sich also entscheiden, ob der Strom vorzugsweise aus positiven oder negativen Trägern besteht. Experimentell findet man in einer Anordnung gemäß Abb. 2.14 für Reingermanium, das mit einer geringen Menge Arsen versehen wurde, negative Teilchen, d. h. Elektronen, als Stromträger. Dotiert man dagegen mit Gallium, so erscheinen die Stromträger positiv. Dieses Ergebnis steht in Widerspruch zur klassischen Elektronentheorie der metallischen Leitung. Wir haben es hier also offenbar mit einer Erscheinung zu tun, die für den Halbleiter typisch ist. Somit kann der HALL-Effekt als das experimentum crucis der neueren Halbleiterphysik bezeichnet werden. Im Gegensatz zu den Metallen hat man es in Halbleitern mit zweierlei Arten von Stromträgern zu tun.

Dieser Sachverhalt muß bei der Ableitung der Formeln für die HALL-Spannung berücksichtigt werden. Vereinfachend sei zunächst angenommen, daß die elektrische Leitfähigkeit nur durch eine Art von Trägern zustande kommt, wobei das Vorzeichen offenbleiben soll.

2.3.3 HALL-Spannung für unipolare Stromleitung. Entsprechend Abb. 2.13 sei $v = v_x$ und $B = B_y$. Die Ladung der Stromträger sei q. Wir fragen nach dem HALL-Feld in z-Richtung F_z. Die LORENTZ-Kraft ist:

$$K = K_z = q[\boldsymbol{v}\ \boldsymbol{B}]_z = q v_x B_y \qquad \text{(Gl. 2.5)}$$

Dieser Kraft K_z wirkt die Kraft des HALL-Feldes $q \cdot F_z$ entgegen. Es muß also sein

$$qF_z + q v_x B_y = 0$$

d. h.

$$-F_z = v_x B_y \qquad \text{(Gl. 2.6)}$$

v_x läßt sich mit Hilfe von Gl. 2.1 durch die Stromdichte j_x ausdrücken, so daß man erhält

$$-F_z = \frac{B_y j_x}{q p} \qquad \text{(Gl. 2.7)}$$

Die HALL-„Konstante“ R_H ist definiert als die Proportionalitätskonstante dieser Gleichung, d. h.

$$F_z = -R_H j_x B_y \qquad \text{(Gl. 2.8)}$$

Es ist also

$$R_H = +\frac{1}{q p} \qquad \text{(Gl. 2.9)}$$

Die HALL-Konstante hat demnach dasselbe Vorzeichen wie q, d. h. wie die Ladungsträger, und ist außerdem um so größer, je kleiner die Trägerdichte p ist.

Nach genauerer Durchrechnung erscheint bei R_H (Gl. 2.9) noch ein Faktor $3\pi/8$. Der bei unserer einfachen Rechnung auftretende Fehler entsteht da-

durch, daß wir die Geschwindigkeit, die in Gl. 2.6 benötigt wird, aus der Gleichung für die Stromdichte, Gl. 2.1, entnommen haben. In Wirklichkeit sind beides Mittelwerte über die einzelnen Richtungen, die sich um den Faktor $3\pi/8$ unterscheiden.

2.3.4 HALL-Spannung für bipolare Stromleitung. Ehe wir die HALL-Spannung für einen bipolaren Leitungsmechanismus ableiten, muß noch ein Versäumnis bei Erwähnung der Gl. 2.2 nachgeholt werden. Beweglichkeiten sind positive Zahlenwerte, bei negativen Ladungen haben aber die Teilchengeschwindigkeit und das elektrische Feld entgegengesetzte Vorzeichen, wir verbessern demnach Gl. 2.2 und schreiben

$$\boldsymbol{v} = \frac{q}{|q|}\mu \boldsymbol{F} \qquad \text{(Gl. 2.10)}$$

Dann gilt diese Beziehung für positive und negative Trägerladungen q.

Besteht der Strom aus positiven Trägern der Dichte p, der Beweglichkeit μ_p und der Ladung $q = +e$ (e ist die Elementarladung), und aus negativen Trägern der Dichte n, der Beweglichtkeit μ_n und der Ladung $q = -e$, so ist die gesamte Stromdichte als die Summe der beiden Teilströme

$$j = j_p + j_n = e p v_p - e n v_n = e(p\mu_p F + n\mu_n F) \qquad \text{(Gl. 2.11)}$$

Während bei der unipolaren Leitung der Strom in z-Richtung null war, ist jetzt lediglich zu fordern, daß die Summe der beiden Teilströme in z-Richtung verschwindet.

Die Kraft in z-Richtung auf ein in x-Richtung strömendes Teilchen ist

$$K_z = qF_z + qv_x B_y = qF'_z \qquad \text{(Gl. 2.12)}$$

mit

$$F'_z = F_z + v_x B_y \qquad \text{(Gl. 2.13)}$$

Wird dies als Feld F unter Berücksichtigung von Gl. 2.10 in Gl. 2.11 eingesetzt, so ergibt sich

$$j_z = e(p\mu_p F_z + p\mu_p^2 F_x B_y + n\mu_n F_z - n\mu_n^2 F_x B_y)$$

Daraus erhält man für $j_z = 0$

$$\frac{F_z}{F_x B_y} = -\frac{p\mu_p^2 - n\mu_n^2}{p\mu_p + n\mu_n} \qquad \text{(Gl. 2.14)}$$

Ersetzt man F_x gemäß Gl. 2.11 durch j_x, so erhält man

$$R_H = -\frac{F_z}{B_y j_x} = \frac{p\mu_p^2 - n\mu_n^2}{e(p\mu_p + n\mu_n)^2} \qquad \text{(Gl. 2.15)}$$

Der Faktor $3\pi/8$ auf der rechten Seite von Gl. 2.15 ist wieder unterdrückt, weil die mittleren Geschwindigkeiten, die in den Strom und die LORENTZ-Kraft eingehen, wieder gleichgesetzt worden sind. Darüber hinaus ist bei Gl. 2.15 vorausgesetzt, daß die durch die magnetischen Ablenkungen hervorgerufenen Dichteunterschiede der Träger zu vernachlässigen sind (näheres darüber siehe z. B. [17]).

Sind die Beweglichkeiten der beiden Trägersorten nicht sehr verschieden, so entscheidet für das Vorzeichen des HALL-Effektes die in der Mehrzahl vorhandene Trägerart. Sobald eine Trägersorte überwiegt, also z. B. $p\mu_p^2 \gg n\mu_n^2$ und $p\mu_p \gg n\mu_n$, erhalten wir wieder die einfache Gl. 2.9.

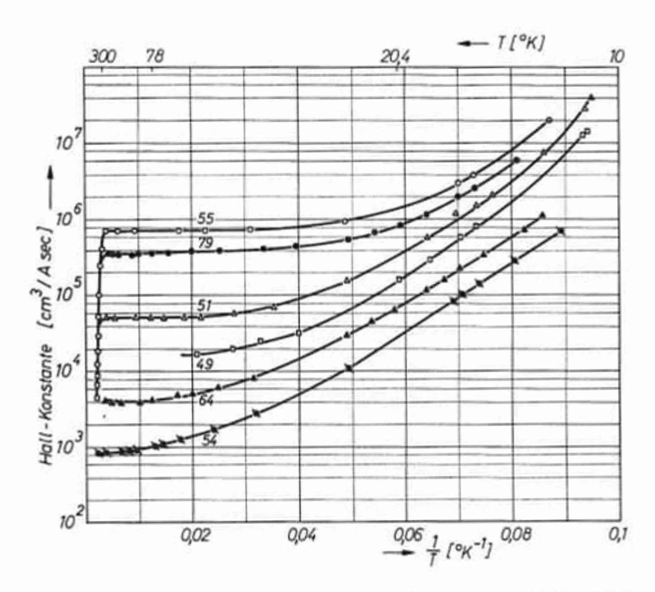


Abb. 2.15. HALL-Konstanten der Proben von Abb. 2.11

In Abb. 2.15 sind die Ergebnisse der HALL-Messungen an den Proben der Abb. 2.11 dargestellt und in Abb. 2.16 die der Proben von Abb. 2.12. Die Vermutung, daß die Trägerdichte im Stör- und Eigenleitungsbereich sich mit der Temperatur stark ändert, ist bestätigt. Für die mit Arsen dotierten Proben ist die HALL-Konstante durchweg negativ, während die Gallium-dotierten Proben im Störleitungsbereich positive HALL-Konstanten haben, die bei dem Übergang in die Eigenleitung zunächst immer kleiner und dann auch negativ werden (gestrichelte Äste).

Der „metallische" Bereich ist durch einen konstanten HALL-Effekt, d. h. eine konstante Trägerdichte, ausgezeichnet, die — nach Gl. 2.9 bestimmt — mit der Störstoffkonzentration übereinstimmt. In diesem Bereich stehen also pro Arsenatom ein Elektron und pro Galliumatom ein positiver Leitfähigkeits-

träger zur Verfügung. Zusammen mit den Leitfähigkeitsmessungen in diesem Bereich und der Formel Gl. 2.3 für die Leitfähigkeit ergibt sich für die Beweglichkeit der Elektronen bei 20°C

$$\mu_n \approx 3900 \frac{\mathrm{cm}^2}{\mathrm{V\,sec}} \qquad \text{(in Germanium)}$$

und für die Beweglichkeit der positiven Stromträger

$$\mu_p \approx 1900 \frac{\mathrm{cm}^2}{\mathrm{V\,sec}} \qquad \text{(in Germanium)}$$

Die Abnahme der Leitfähigkeit mit steigender Temperatur im metallischen Bereich rührt von einer Abnahme der Beweglichkeit her, die übrigens in derselben Weise erfolgt wie bei Metallen. In Anlehnung an die Verhältnisse in Metallen — konstante Trägerdichte, wachsende Behinderung der Elektronen durch die Zusammenstöße mit dem Gitter — besteht demnach unsere Bezeichnung dieses Bereiches konstanter Trägerdichte zu Recht.

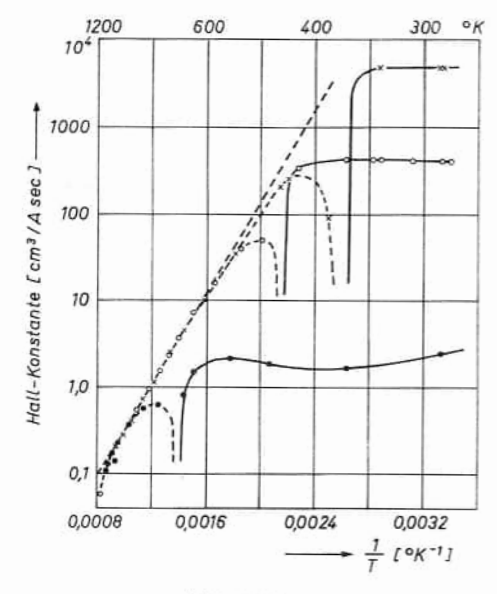
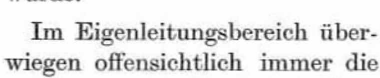


Abb. 2.16.
HALL-Konstanten der Proben von Abb. 2.12
Die gestrichelten Teile sind negativ.

Die Abnahme der Leitfähigkeit nach tiefen Temperaturen hin (Abb. 2.10, 2.11) muß von einer Abnahme der Trägerdichte herrühren, wie aus der Zunahme der HALL-Konstanten (Abb. 2.15) hervorgeht. Die Beweglichkeit der Träger mit abnehmender Temperatur wird sogar größer. Dies ist in Abb. 2.17 dargestellt, die aus Abb. 2.11 und Abb. 2.15 gewonnen wurde.



Im Eigenleitungsbereich überwiegen offensichtlich immer die Elektronen. Der Verlauf der HALL-Konstanten für Werte, die klein gegenüber dem Wert bei Störleitung sind, ist mit Gl. 2.15 verträglich, wenn man annimmt, daß in diesem Bereich die Dichte der positiven und negativen Träger gleich ist. In diesem Falle erhält man aus Gl. 2.15 mit $p = n$

$$R_H = \frac{1}{en} \frac{\mu_p - \mu_n}{\mu_p + \mu_n} \qquad \text{(Gl. 2.16)}$$

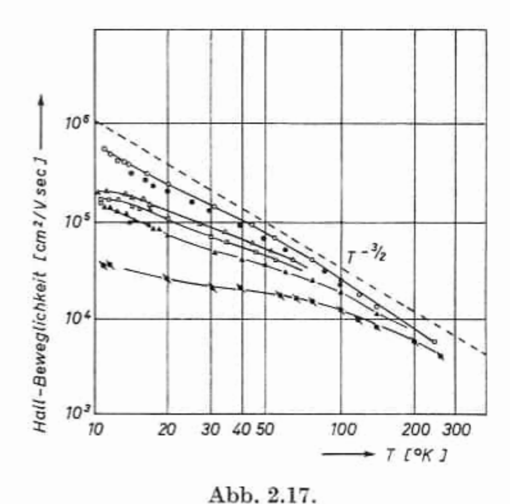


Abb. 2.17.
Elektronenbeweglichkeit, berechnet aus Abb. 2.11 und 2.15

Da $\mu_n > \mu_p$ ist, ist der negative Wert der HALL-Konstanten in diesem Bereich verständlich. Im Bereich der Eigenleitung nimmt der Betrag der HALL-Konstanten stark ab, d. h. die Trägerdichte stark zu.

Zusammenfassend läßt sich aus den Widerstands- und HALL-Effekt-Messungen also schließen:

1. Im Störleitungsbereich nimmt bei tiefer Temperatur die Trägerdichte, die von den Störatomen herrührt, mit der Temperatur zunächst stark zu, um zwischen 100 und 300°K praktisch konstant zu bleiben. In diesem Bereich ist $n = N_D$ bzw. $p = N_A$, wenn N_D die As-Dichte und N_A die Ga-Dichte im Germanium darstellt.
2. Im Eigenleitungsbereich findet man ebenfalls eine starke Zunahme der Trägerdichte mit der Temperatur. Hier ist die Elektronendichte und die Dichte der positiven Träger gleich.
3. Der Temperaturverlauf der Beweglichkeit ist von demjenigen der Elektronen in Metallen nicht wesentlich verschieden.

2.4 Zusammenfassung der bisher erläuterten Experimente in einer Modellvorstellung

2.4.1 Eigenleitung. Die Eigenleitung ist offenbar eine Eigenschaft des Germaniumkristalls, die allenfalls bei tieferen Temperaturen von der Störleitung teilweise überdeckt sein kann. Bei genügend hohen Temperaturen ist sie auch bei stark fremdstoffhaltigen Proben immer an dem starken Anstieg der Leitfähigkeit mit der Temperatur zu erkennen. Zur Erklärung der Eigenleitung bei Germanium läßt sich folgendes sagen:

Reines Germanium ist bei tiefer Temperatur ein elektrischer Isolator. Die einzelnen Germaniumatome sind — da Germanium vierwertig ist — tetraedrisch aneinander gebunden; in Abb. 2.18 wird eine zweidimensionale Darstellung der Bindungsverhältnisse gegeben. Die Doppelstriche zwischen den Germaniumatomen sollen je zwei Valenzelektronen darstellen, die, ebenso wie bei der Wasserstoffmolekülbindung, gleichzeitig zwei benachbarte Germaniumrümpfe umkreisen und die Bindungen bewerkstelligen. Diese Verhältnisse gelten

für sehr tiefe Temperaturen, eine elektrische Leitfähigkeit ist nicht vorhanden (Teil a).

Bei höherer Temperatur hat die Energie der Gitterschwingungen einige wenige dieser Bindungen gelöst. Die befreiten Elektronen, die zuvor für Bindungsaufgaben eingesetzt waren, sind, soweit sie sich an Stellen des Kristalls befinden, die mit Valenzelektronen versehen sind, als frei zu betrachten. Ein elektrisches Feld bewegt diese Elektronen in ähnlicher Weise wie die Elektronen eines Metalls (Teil b).

=Ge=Ge=Ge=Ge=Ge=
=Ge=Ge=Ge=Ge=Ge=
=Ge=Ge=Ge⁺–Ge=Ge=
=Ge=Ge=Ge⁺–Ge=Ge=
B A
$\vec{F}$
a b

Abb. 2.18. **Valenzgittermodell** und Eigenleitung.

Neu und charakteristisch für Halbleiter ist nun die Tatsache, daß auch die Stellen, an denen ein Elektron abgelöst worden ist, in einem elektrischen Feld verschoben werden können. Es ist nämlich möglich, daß unter der Wirkung des elektrischen Feldes die Bindungslücke an der Stelle A der Abb. 2.18 von einem Valenzelektron aus der Nachbarschaft — z. B. aus der Stelle B — geschlossen wird. Damit wandert die Stelle im Kristall, die einen Valenzelektronendefekt hat, nach links. Da die mit Valenzelektronen abgesättigten Gitterplätze neutral sind, erzeugt ein Valenzelektronendefekt eine positive Ladung bei der davon betroffenen Gitterstelle. Wandert die „defekte Stelle“, so wandert auch die positive Ladung. Man sieht, daß Platzwechsel von Valenzelektronen, die vorzugsweise in einer Richtung stattfinden, als eine Wanderung positiver Teilchen in der Gegenrichtung erscheinen. Die Trägheit der Bewegungen dieser positiven Ladungen ist nicht viel größer als diejenige der Elektronen, da es sich im Grunde um eine Elektronenbewegung handelt.

Es hat sich als zweckmäßig erwiesen, die Bewegungen der positiven Stellen mit Valenzelektronendefekten durch ein positiv geladenes Teilchen — das „Defektelektron“ oder kurz „Loch“ — zu beschreiben. Entsprechend seinem Verhalten im elektrischen Feld wird es — ebenso wie die freien Elektronen — durch eine positive Masse — die Löchermasse — und eine Beweglichkeit — die Löcherbeweglichkeit — charakterisiert. Bei genauer Betrachtung des Platzwechselmechanismus von Valenzelektronen, der schließlich zur Bewegung eines einzigen Defektelektrons durch einen Germaniumkristall führt, fällt auf, daß — im Gegensatz zu den freien Elektronen — die Wanderung des Defektelektrons nichts anderes ist als die Summe der einzelnen Sprünge der Valenzelektronen. Dieser Sachverhalt wird in Kapitel 3 quantentheoretisch unterbaut werden. Man kann also sagen, daß mit der Vorstellung der Defektelektronen zweckmäßig und anschaulich die Gesamtheit derjenigen Elektronen im Festkörper beschrieben wird, die sich in Bindungszuständen befinden. Die aus solchen Bindungszuständen befreiten Elektronen gehören zu der davon verschiedenen Klasse der freien Elektronen und sind, wie bei den Metallen, mit negativer Ladung und positiver Elektronenmasse zu beschreiben.

In einem ungestörten Halbleiter werden demnach durch die Wechselwirkung der thermischen Gitterschwingungen (Schallquanten oder Phononen) mit den Gitteratomen Elektronen und Defektelektronen in gleicher Anzahl geschaffen. Diesen Prozeß nennt man Generation von Paaren von Elektronen und Defektelektronen. Die Wahrscheinlichkeit der Lösung eines Valenzelektrons hängt von der Zahl der gebundenen Valenzelektronen ab, die so lange als konstant zu betrachten ist, als wenig Valenzelektronen befreit werden. Das ist immer der Fall. Außerdem ist sie aber abhängig von der Zahl der Gitteratome, die Schwingungsenergien besitzen von mindestens der Größe der Loslösungsenergie E_G eines Valenzelektrons. Die Wahrscheinlichkeit, eine bestimmte Schwingungsenergie E bei einem Gitteratom zu finden, ist nach dem BOLTZMANNschen Satz proportional $e^{-E/kT}$, wo T die absolute Temperatur des Gitters und k die BOLTZMANNsche Konstante ist. Die Wahrscheinlichkeit für alle Schwingungsenergien größer als die Bindungsenergie der Valenzelektronen $E > E_G$ ist proportional $kT\, e^{-E_G/kT}$. Die Generation nimmt also für $kT \ll E_G$ etwa exponentiell mit der Temperatur zu.

Außer der Generation von Elektron und Loch findet auch der rückläufige Prozeß statt, der darin besteht, daß ein freies Elektron wieder in eine Valenzelektronenlücke zurückspringt (Rekombination). Dabei verschwinden gleichzeitig ein Elektron und ein Defektelektron. Am einfachsten ist dabei die Annahme, daß die Rekombination proportional der Dichte n der Elektronen und der Löcherdichte p ist. Dieser Prozeß begrenzt also die durch Generation erzeugte Teilchendichte eines erwärmten Gitters insofern, als die Rekombination mit wachsender Dichte zunimmt. Im thermischen Gleichgewicht, wo Generation und Rekombination gleich häufig sind, hat man demnach zu erwarten

$$n\,p \sim e^{-E_G/kT} \qquad \text{(Gl. 2.17)}$$

(kT auf der rechten Seite von Gl. 2.17 wurde als unwesentlich gegenüber der Exponentialfunktion unterdrückt.) Bei Eigenleitung ist $n = p$. Wir nennen diese Trägerdichte Eigenleitungsdichte n_i und können schreiben:

$$n_i = n = p \sim e^{-E_G/2kT} \qquad \text{(Gl. 2.18)}$$

Damit im Einklang ist der Verlauf von Widerstand und HALL-Effekt im Eigenleitungsbereich der in Abschnitt 2.3.4 gemessenen Proben, die entsprechend Gl. 2.18 in Abb. 2.11, 2.12, 2.15 und 2.16 gegen $1/T$ statt gegen T aufgetragen wurden. Die starke Zunahme der Leitfähigkeit mit der Temperatur im Eigenleitungsbereich ist eine Folge der starken Zunahme von Elektronen- und Löcherdichte durch die exponentiell mit $-1/T$ veränderliche Generation.

Aus Abb. 2.14 und 2.15, der Formel Gl. 2.16 für den HALL-Koeffizienten im Eigenleitungsbereich und der in Gl. 2.18 angegebenen Temperaturabhängigkeit der Eigenleitungsträgerdichte erhält man für die Bindungsenergie der Valenzelektronen im Germaniumgitter einen Wert von 0,78 eV. (Zu einer

genauen quantitativen Rechnung sind allerdings noch temperaturabhängige Faktoren in Gl. 2.18 und in Gl. 2.16 zu berücksichtigen, worauf wir hier nicht eingehen können. Wir verweisen dazu auf die Arbeit von MORIN und MAITA [18], der auch der Zahlenwert für E_G entnommen ist.) Das Dichteprodukt von Elektronen- und Löcherdichte ergibt sich zu $6 \cdot 10^{26}$ cm^{-6} bei Zimmertemperatur und verändert sich mit der Temperatur gemäß Gl. 2.17 etwa exponentiell.

Bei der Betrachtung des Wechselspiels von Generation und Rekombination ist die Frage nach der mittleren Lebensdauer eines Elektrons bzw. Defektelektrons zwischen Generation und Rekombination interessant. Es zeigt sich später, daß diese Größe für Gleichrichter- und Transistoreffekt von großer Bedeutung ist. Hier mag die Bemerkung genügen, daß die Bilanz, die zu Gl. 2.18 führte, unabhängig von dieser Trägerlebensdauer ist. Ist irgendwo in einem Kristall die Trägerlebensdauer kleiner (z. B. in der Oberfläche), so führt das nach einem Satz der Statistik (Prinzip des detaillierten Gleichgewichts) auch zu erhöhter Generation.

Zusammenfassend lassen sich die Vorgänge bei Eigenleitung durch die folgende Reaktionsgleichung beschreiben:

$$\text{Valenzelektron} \underset{\text{Rekombination}}{\overset{\text{Generation}}{\rightleftarrows}} \text{freies Elektron} + \text{freies Defektelektron}$$

2.4.2 Störleitung.

2.4.2.1 Elektronenleitung. Durch geringe Arsenzusätze zu Reingermanium erhält man, wie bereits in Abschnitt 2.3.4 erwähnt, bei Zimmertemperatur elektronenleitendes Germanium mit einer Elektronendichte, die gleich der Dichte der eingebauten Arsenatome ist. Das ist zu verstehen, wenn man annimmt, daß das Arsen atomdispers im Germanium verteilt ist und jede so geschaffene Störstelle ein freies Elektron liefert.

Ein Fremdatom kann grundsätzlich auf zwei verschiedene Arten in einen Festkörper eingebaut werden. Entweder sitzt der Fremdstoff auf den Gitterplätzen des Wirtsgitters (d. h. in unserem Fall, daß Arsenatome Plätze einnehmen die im ungestörten Gitter von Germaniumatomen besetzt sind), oder das Fremdatom drängt sich an einem oder einigen dafür geeigneten Orten zwischen die Gitteratome.

Die Elemente der III. und V. Gruppe des Periodischen Systems, die man vorzugsweise zum Dotieren von Germanium verwendet, werden normalerweise auf Gitterplätzen des Germaniumgitters eingebaut. Das ergibt sich aus den mit Röntgenstrahlen durchgeführten Messungen der Gitterkonstanten bei verschiedenen Störstoffkonzentrationen.

Das Arsenatom ist fünfwertig, hat also ein Valenzelektron mehr als das Ge-Atom, dessen Stelle es im Gitter einnimmt. Die Umgebung einer Arsen-

```
 ||    ||    ||    ||    ||    ||    ||
=Ge = Ge = Ge = Ge = Ge = Ge = Ge =
 ||    ||    ||    ||    ||    ||    ||
=Ge = As = Ge = Ge = As = Ge = Ge =
 ||    ||    ||    ||    ||    ||    ||
=Ge = Ge = Ge = Ge = Ge = Ge = Ge =
 ||    ||    ||    ||    ||    ||    ||
       a                 b
```

Abb. 2.19.
Valenzgittermodell mit Donatoren.

störstelle kann demnach durch Abb. 2.19a veranschaulicht werden. Vier von den fünf Valenzelektronen des Arsens übernehmen die Bindungsaufgaben dieses Gitterplatzes zu den Germaniumnachbarn. Das 5. Elektron ist infolge der hohen Dielektrizitätskonstanten des Germaniums sehr lose gebunden. Bereits bei Zimmertemperatur sind deshalb infolge der thermischen Gitterschwingungen sämtliche Arsenatome ionisiert (vgl. Abb. 2.19b). Die abgetrennten Elektronen sind im Gitter frei beweglich und stehen als Leitfähigkeitselektronen zur Verfügung, während die positiven Arsenionen fest sind und (im Gegensatz zu den Defektelektronen in Abschnitt 2.4.1) im elektrischen Feld nicht bewegt werden. Nur bei sehr hohen Temperaturen ist eine Wanderung des Arsenions im Ge-Gitter möglich. Eine solche Ionenleitfähigkeit ist aber auf alle Fälle außerordentlich klein, so daß wir sie nicht zu berücksichtigen brauchen.

Da fünfwertige Atome, wie Arsen, Antimon, Phosphor, Wismut, als Elektronenspender wirken, bezeichnet man sie als *Donatoren*. Der mit solchen Stoffen dotierte Halbleiter weist Elektronen, also negative Leitfähigkeitsträger, auf; man bezeichnet ihn also deshalb oft kurz als n-Typ-Halbleiter. Die „Reaktion“, die neben der Eigenleitung durch Zugabe eines Donators im Halbleiter abläuft, kann also folgendermaßen geschrieben werden:

$$\underset{\text{ortsfest}}{\text{As}} \rightarrow \underset{\text{ortsfest}}{\overset{+}{\text{As}}} + \underset{\text{beweglich}}{\text{Elektron } \ominus}$$

bzw. allgemein $\quad \text{Donator} \rightarrow \overset{+}{\text{Donator}} + \ominus$

Im reinen Germanium ist bei Zimmertemperatur infolge der Eigenleitung eine Dichte von etwa $2 \cdot 10^{13}$ Elektronen (und Löchern) pro cm^3 vorhanden. Durch Zulegieren von Arsen zu Germanium erhält man eine zusätzliche Elektronendichte, die ungefähr der Arsenkonzentration entspricht. Die Arsenverunreinigung macht sich also schon bei einer Konzentration von $10^{13}/cm^3$ bemerkbar. Da sich im Germaniumgitter etwa 10^{23} Germaniumatome pro cm^3 befinden, entspricht dies einer relativen Konzentration von nur 10^{-10} Atomteilen Arsen. Andere Stoffe dotieren ähnlich wie Arsen, und es ist somit verständlich, daß nur Germanium von der außerordentlichen Reinheit von weniger als ein Störatom auf 10^{10} Germaniumatome bei Zimmertemperatur die Eigenleitung in Erscheinung treten läßt.

Andererseits steigt die Eigenleitung exponentiell mit der Temperatur an — bei 200°C ist die Elektronen- und Löcherdichte gemäß Tabelle 1 bereits 10^{16} pro cm^3; bei genügend hoher Temperatur übersteigt die Eigenleitfähigkeit immer die Wirkung eines verhältnismäßig hohen Störstoffgehaltes.

Tabelle 1
Eigenleitungsdichten von Ge, Si und Te

T	Germanium ($E_G = 0{,}78$ eV)	Silizium ($E_G = 1{,}2$ eV)	Tellur ($E_G = 0{,}37$ eV)
20 °C	$1{,}3 \cdot 10^{13}$ cm^{-3}	$7{,}7 \cdot 10^{9}$ cm^{-3}	$5 \cdot 10^{16}$ cm^{-3}
50 °C	$6{,}6 \cdot 10^{13}$ cm^{-3}	$8{,}4 \cdot 10^{10}$ cm^{-3}	$8 \cdot 10^{16}$ cm^{-3}
100 °C	$5{,}6 \cdot 10^{14}$ cm^{-3}	$1{,}9 \cdot 10^{12}$ cm^{-3}	$1{,}9 \cdot 10^{17}$ cm^{-3}
150 °C	$2{,}9 \cdot 10^{15}$ cm^{-3}	$2{,}2 \cdot 10^{13}$ cm^{-3}	$4{,}5 \cdot 10^{17}$ cm^{-3}
200 °C	$1{,}1 \cdot 10^{16}$ cm^{-3}	$1{,}5 \cdot 10^{14}$ cm^{-3}	$7{,}5 \cdot 10^{17}$ cm^{-3}
250 °C	$3{,}2 \cdot 10^{16}$ cm^{-3}	$7{,}0 \cdot 10^{14}$ cm^{-3}	$1{,}1 \cdot 10^{18}$ cm^{-3}

Silizium hat bei Zimmertemperatur eine Eigenleitungsdichte von weniger als 10^{10} Träger/cm^3 ($E_G = 1{,}2$ eV), was einem spezifischen Widerstand von etwa 300000 Ohm · cm entspricht. Bei seiner höheren Schmelztemperatur (1420°C) und der damit verbundenen größeren Möglichkeit der Verunreinigung ist es verständlich, daß es erst neuerdings gelungen ist, einkristallines Silizium herzustellen, das bei Zimmertemperatur einwandfrei eigenleitend ist. Der dazu gehörende Störstoffgehalt muß unter 10^{-13} Atomteilen liegen. Dies kann nur durch Elimination des Bors auf chemischem Wege und anschließendes tiegelfreies Zonenschmelzen erreicht werden. Derartige Reinheit kann man auch bei Germanium nur auf diese Weise erhalten [19].

Als drittes Beispiel ist in Tab. 1 Tellur aufgeführt, das nach der Arbeit von Kronmüller, Jaumann und Seiler berechnet wurde [20]. Die Aktivierungsenergie der Valenzelektronen des Tellurs beträgt nur etwa 0,37 eV, was zur Folge hat, daß bereits ein unter Halbleitergesichtspunkten störstofffreies Tellur bei Zimmertemperatur eigenleitend ist. Will man hier Störstoffkonzentrationen von einigen 10^{15} bis 10^{16} pro cm^3, wie sie für den Halbleiterphysiker interessant sind, feststellen, so hat man Hall-Effekt und Leitfähigkeit etwa bei der Temperatur der flüssigen Luft zu messen.

2.4.2.2 Löcherleitung. Bereits in Abschnitt 2.3.2 wurde festgestellt, daß Gallium das Germanium so dotiert, daß sich ein positiver Hall-Koeffizient ergibt (positive Leitfähigkeitsträger). Stromträgerdichte und Störstoffdichte sind bei Zimmertemperatur wieder gleich, und die Beweglichkeit ist etwa 50% der Elektronenbeweglichkeit. Auch in die Gl. 2.16 wurde mit Erfolg die bei Gallium-dotierten Proben erhaltene Trägerbeweglichkeit eingesetzt. Wir können deshalb annehmen, daß durch Dotieren mit Gallium — oder allgemein mit Stoffen der III. Gruppe des Periodischen Systems — dieselben Defektelektronen entstehen wie bei der Eigenleitung.

Die Umgebung einer Galliumstörstelle im Germaniumgitter ist in Abb. 2.20a skizziert. Die drei Valenzelektronen des Galliums ersetzen drei der vier Valenzelektronen des Germaniums, das im ungestörten Gitter an Stelle des Galliums sitzt. Das vierte Bindungselektron, welches von den das

```
   ‖     ‖     ‖     ‖     ‖+    ‖     ‖
= Ge = Ge = Ge = Ge = Ge = Ge = Ge =
   ‖     ‖     ‖     ‖     |     ‖     ‖
= Ge =[Ga]- Ge = Ge = Ge =[Ga]= Ge =
   ‖     ‖     ‖     ‖     ‖     ‖     ‖
= Ge = Ge = Ge = Ge = Ge = Ge = Ge =
   ‖     ‖     ‖     ‖     ‖     ‖     ‖
         a                 b
```

Abb. 2.20.
Valenzgittermodell mit Akzeptoren.

Gallium umgebenden Germaniumatomen gefordert wird, fehlt. Es besteht aber die Möglichkeit, daß gemäß Abb. 2.20 b ein Valenzelektron eines der Germaniumnachbarn als viertes Bindungselektron einspringt. Das Gallium, das zuvor neutral war, ist jetzt negativ geladen, und das Germaniumatom, das ein Valenzelektron abgegeben hat, trägt eine positive Ladung. Das fehlende Valenzelektron des Germaniumatoms kann nun in derselben Weise wieder von Germaniumnachbarn ersetzt werden wie vorher bei der Eigenleitung beschrieben wurde, d. h. durch Dotieren mit Gallium haben wir so viel Löcher geschaffen wie Galliumatome im Germanium enthalten sind. Die Löcher sind frei beweglich, die negativ geladenen Galliumionen dagegen wieder unbeweglich. Da Gallium — allgemein die Elemente der III. Gruppe — Elektronen dem Germaniumgitter entzieht, gehört es zur Gruppe der Akzeptoren. Der Halbleiter ist p-leitend, d. h. seine Stromträger haben positive Ladung. Als Akzeptoren wirken also die Elemente Bor, Aluminium, Gallium, Indium und Thallium.

Die entsprechende „Reaktionsgleichung" kann geschrieben werden

$$\underset{\text{ortsfest}}{\mathrm{Ga}} \;\rightarrow\; \underset{\text{ortsfest}}{\mathrm{Ga}^-} \;+\; \underset{\text{beweglich}}{\text{Defektelektron}\ \oplus}$$

oder allgemein $\quad \text{Akzeptor} \rightarrow \text{Akzeptor}^- + \text{Defektelektron}\ \oplus$

2.4.2.3 Aktivierungsenergie von Donatoren und Akzeptoren. Aus der geringen Veränderung der Hall-Konstanten im Störleitungsbereich (vgl. Abb. 2.15 und 2.16) geht hervor, daß die Bindungsenergie der Leitfähigkeitselektronen an die Donatoren außerordentlich gering sein muß. Über die genaue Größe können nun Hall-Messungen und überschlägig Leitfähigkeitsmessungen bei Temperaturen unterhalb 50°K Auskunft geben, wie sie in Abb. 2.11 und 2.15 wiedergegeben sind.

Um aus ihnen die Bindungsenergie der Leitfähigkeitselektronen an die Arsenatome zu entnehmen, haben wir eine ähnliche Überlegung anzustellen wie in Abschnitt 2.4.1. Von N_D Donatoren pro cm³ seien n ionisiert, d. h. es seien n Leitfähigkeitselektronen pro cm³ vorhanden. Diese Elektronendichte n kann aus der Bilanz der Generation der Elektronen aus neutralen Donatoren und der Rekombination der Leitfähigkeitselektronen mit den bereits ionisierten Donatoren ermittelt werden. Die erste Zahl ist im wesentlichen ein Produkt von einer Größe $e^{-E_D/kT}$, d. h. der Wahrscheinlichkeit für thermische Stöße mit Energien $E > E_D$ (E_D ist die Bindungsenergie der Leitfähigkeitselektronen an die Donatoren) und der Dichte der noch neutralen Donatoren $N_D - n$. Die

Generationsrate ist also proportional $(N_D - n) \cdot e^{-E_D/kT}$. Die Rekombination der Leitfähigkeitselektronen ist proportional ihrer Dichte n und der Dichte der ionisierten Donatoren (ebenfalls $= n$), also proportional n^2. Es ist dann im Gleichgewicht: Generation gleich Rekombination, also

$$n^2 = \text{const} \cdot (N_D - n)\, e^{-E_D/kT} \qquad \text{(Gl. 2.19)}$$

Dasselbe Ergebnis erhalten wir, wenn wir das Massenwirkungsgesetz auf die Reaktionsgleichung der Donatoren (Abschnitt 2.4.2.1) anwenden. Die Massenwirkungskonstante ist proportional $e^{-E_D/kT}$.

Es lassen sich nun leicht zwei Grenzfälle aus Gl. 2.19 erkennen:

1. Schwache Ionisation ($n \ll N_D$): Die Dichte der Leitfähigkeitselektronen ist klein gegenüber der Störstellendichte und proportional der Wurzel aus der Störstellendichte:

$$n \sim \sqrt{N_D}\, e^{-E_D/2kT} \qquad \text{(Gl. 2.20)}$$

Dieser Fall liegt gemäß Gl. 2.19 bei tiefen Temperaturen vor. Die Bindungsenergie der Leitfähigkeitselektronen an die Donatoren ergibt sich daraus mit Messungen wie Abb. 2.15 zu

$$E_D^{(As)} = 0{,}013 \text{ eV}$$

2. Starke Ionisation: Sie ergibt sich nach Gl. 2.19 für $kT \gg E_D$. Da die Konstante in Gl. 2.19 etwa gleich $5 \cdot 10^{18}\,\text{cm}^{-3}$ ist (vgl. Abschnitt 4.1.3) und im allgemeinen $n < 10^{17}\,\text{cm}^{-3}$, wird die Elektronendichte gleich der Störstellendichte, wie wir es im Germanium bei Temperaturen über etwa 100°K feststellen, also $n \approx N_D$.

Im Fall schwacher Ionisation sind nicht alle Störstellen ionisiert. Durch Einstrahlung von Licht oder durch Elektronenstoß lassen sich z. B. noch zusätzliche Stromträger gewinnen, weshalb man diesen Fall auch als „Reserve“ bezeichnet. Bei starker Ionisation sind praktisch alle Störstellen frei von Stromträgern. Der Halbleiter befindet sich im Zustand der „Erschöpfung“.

Die im vorliegenden Abschnitt am Fall des n-Typs abgeleiteten Begriffe und Gleichungen lassen sich ohne weiteres auf den p-Typ, d. h. auf das Gleichgewicht zwischen Akzeptoren und Löchern, übertragen.

2.4.2.4 Kompensation. Sind Donatoren und Akzeptoren in gleicher Dichte vorhanden, so kann man aus den oben entwickelten Vorstellungen zur Störleitung erwarten, daß gemäß Abb. 2.19 und 2.20 das überzählige fünfte Elektron des Donators, das bei Zimmertemperatur frei beweglich ist, die durch den Akzeptor geschaffene Bindungslücke eines Germaniumatomes ausfüllt oder direkt das vierte Bindungselektron an der Akzeptorstelle liefert. Das Ergebnis ist in beiden Fällen, daß kein freier Stromträger vorhanden ist, sondern nur positive Donatorionen und negative Akzeptorionen.

Sind Donatoren und Akzeptoren in verschiedener Dichte vorhanden, so bestimmen die im Überschuß vorhandenen Störstoffe die Trägersorte. Mehr Donatoren als Akzeptoren liefert Elektronenleitung, mehr Akzeptoren als Donatoren Löcherleitung. Quantitativ haben wir also bei Zimmertemperatur nicht mehr $n = N_D$ bzw. $p = N_A$, sondern $n = N_D - N_A$ bzw. $p = N_A - N_D$. Teilweise kompensierte Halbleiter weisen geringere Beweglichkeit auf als unkompensierte gleicher Trägerdichte. Die elektrisch geladenen Störstoffionen lenken die Stromträger auf ihrer Bahn durch den Kristall mehr als die entsprechenden Germaniumatome ab, so daß eine geringere Beweglichkeit resultiert. Auch das Tieftemperaturverhalten weicht von demjenigen unkompensierter Dotierung ab.

2.4.3 Termschema des Halbleiters (Bändermodell). Für jedes Halbleitermaterial läßt sich ein Schema aufstellen, in dem die Energien der verschiedenen Möglichkeiten einer Elektronenbindung übersichtlich zusammengefaßt sind. Auf der Ordinate der Abb. 2.21 wurde die Energie der Elektronen in den verschiedenen möglichen Bindungszuständen aufgetragen. Als Einheit diene z. B. das Elektronenvolt (eV), d. h. das Produkt der Elementarladung mit der Spannungseinheit. Ein eV entspricht $1{,}6 \cdot 10^{-19}$ Wattsekunden. Die Abszisse zeigt irgendeine Lineardimension des Halbleiterkristalls. Die Energie null habe ein Elektron, das genügend weit vom Halbleiter entfernt ist, so daß alle Elektronen im Halbleitergitter negative Energien besitzen.

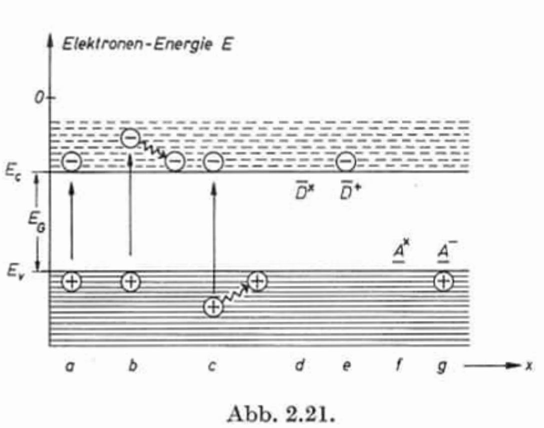


Abb. 2.21.
Bändermodell.

2.4.3.1 Bändermodell bei Eigenleitung. Die Abtrennarbeit für das erste der vier Bindungselektronen sei $|E_V|$, d. h. es muß die Energie $|E_V|$ aufgebracht werden, um ein Valenzelektron aus dem Kristallverband zu entfernen. Im Gegensatz zur Anschauung haben aber nicht alle Bindungselektronen dieselbe Energie E_V. Nach dem Pauli-Prinzip können im ganzen Kristall nur zwei Elektronen, entsprechend zweier verschiedener Spineinstellungen, dieselbe Bindungsenergie E_V haben. Die Bindungsenergien aller anderen Valenzelektronen sind größer, ihre Energieterme liegen also tiefer als E_V. Der Abstand der einzelnen Terme ist aber kleiner als 10^{-20} eV, so daß die verschiedenen Energien der Bindungselektronen in unserem Energieschema (siehe

Abb. 2.21) als ein *Energieband* erscheinen. Die Struktur des Bandes der Valenzelektronen, oder kurz des Valenzbandes, sei im Augenblick nicht betrachtet, wichtig ist lediglich die obere und für manche Fälle die untere Grenze. Wird von der Kristalloberfläche abgesehen, so gilt das eben gesagte für alle Bindungselektronen des Kristalls. Die einzelnen Elektronen können wegen ihrer Wellennatur nicht lokalisiert werden; auch ihre höchsten Energiezustände liegen sehr dicht beisammen. Wir dürfen deshalb die obere Grenze des Valenzbandes mit einer durchgehenden Linie im Abstand E_V vom Nullniveau kennzeichnen.

Wird durch die thermische Energie des Gitters, Einstrahlung von Licht oder dergleichen ein Elektron aus dem Valenzband herausgelöst (nicht aber aus dem Kristall entfernt), so benötigt man dazu bei Ge, wie wir im Abschnitt 2.4.1 gezeigt haben, etwa die Energie $E_G = 0{,}78$ eV. Das geschaffene Leitungselektron besitzt an allen Stellen des Kristalls diese Energie. Wir können also für das Leitungselektron in Abb. 2.21 einen weiteren Energieterm $E_C = E_V + E_G$ durchzeichnen. Da die Leitungselektronen trotz ihrer freien Beweglichkeit immer noch im Kristall energetisch festgehaltene Elektronen sind, ist die Energie E_C noch negativ. Auch für die Leitungselektronen gilt das PAULI-Prinzip, d. h. auch hier liegt ein Band von Energietermen vor. E_C ist die kleinste Energie oder der untere Rand des Leitungsbandes. Da die Dichte der Leitungselektronen im Halbleiter sehr gering und die Differenz der einzelnen Energieterme auch im Leitungsband etwa so klein wie im Valenzband ist, besetzen die Leitungselektronen praktisch nur den unteren Rand des Leitungsbandes, sie haben also alle nahezu dieselbe Energie E_C.

Zwischen dem oberen Rand des Valenzbandes und dem unteren des Leitungsbandes sind beim ungestörten Halbleiter keine stationären Elektronenzustände möglich. Entweder ist ein Elektron im Valenzband oder, falls ihm mindestens die Aktivierungsenergie $E_G = E_C - E_V$ zugeführt wird, im Leitungsband, aber nie liegt die Energie eines Elektrons für längere Zeit zwischen beiden. Diese Energielücke oder Bandlücke entspricht dem Abstand zwischen zwei diskreten Energietermen im einzelnen, isolierten Atom, z. B. dem Wasserstoffatom. Entweder ist das Wasserstoffelektron in der einen oder in der anderen Energiestufe, aber nie dazwischen, d. h. nicht länger als 10^{-8} sec.

Ein losgelöstes Valenzelektron ist anschaulich in Abb. 2.21a skizziert: Ein Leitungselektron befindet sich auf der Energie E_C, im Valenzband ist ein positiv geladenes Loch entstanden. Falls das Leitfähigkeitselektron mit einer Energie größer als E_G abgetrennt worden ist (Abb. 2.21b), verliert es durch Zusammenstöße mit dem Gitter so lange Energie, bis es am unteren Rand des Leitfähigkeitsbandes ist, d. h. seine Energie E_C beträgt. Dies geht innerhalb sehr kurzer Zeiten (etwa 10^{-10} sec) vor sich. Auf dem Niveau E_C bleibt es längere Zeit; es sinkt zunächst in der Energieskala nicht mehr tiefer. Jedoch besitzt es dann über längere Zeiten hin (z. B. 10^{-6} sec) eine gewisse Wahrscheinlichkeit, wieder in ein Loch im Valenzband zurückzufallen, d. h. mit einem Loch zu rekombinieren. Wie in Abschnitt 2.4.1 ausgeführt, sind im thermodynamischen

Gleichgewicht Generationsrate und Rekombinationsrate gleichgroß, so daß sich bei einer bestimmten Temperatur immer eine bestimmte Anzahl von Leitungselektronen und Defektelektronen einstellt.

Die fast freie Beweglichkeit der Elektronen und Löcher ist durch die durchgezogenen bzw. gestrichelten Linien in Höhe der Energieterme in Abb. 2.21 veranschaulicht.

Stehen zur Anregung der Valenzelektronen Energien größer als E_G zur Verfügung, so können die Leitfähigkeitselektronen auch von tiefer als E_V liegenden Valenzelektronentermen gebildet werden (Abb. 2.21c). Dadurch entsteht ein unbesetzter Elektronenenergieterm mit einer Energie kleiner als E_V. Die Valenzelektronen mit höheren Energielagen fallen so lange in das energetisch tiefer gelegene Loch, bis dieses gleichsam wie eine Gasblase im See der Valenzelektronen an der Oberfläche ist, d. h. ein Elektronenzustand nahe E_V unbesetzt ist.

2.4.3.2 Bändermodell bei Störleitung. Die Bindungsenergie der Leitfähigkeitselektronen an die ionisierten Donatoren der V. Gruppe beträgt bei Ge etwa 0,01 eV. Ist ein Elektron an einen Donator gebunden, so ist seine Energie um diesen Betrag niedriger als am unteren Rand des Leitfähigkeitsbandes. Die Energie der Elektronen in den Donatoren oder kurz die Energie der Störterme, hat demnach die in Abb. 2.21d und 2.21e eingetragene Lage. $D^\times$ bedeutet, daß das Elektron im Störterm, der Donator also neutral ist (Fall *d*). D^+ stellt einen ionisierten Donator dar, d. h. einen solchen, der sein Elektron verloren hat (Fall *e*). Bei der außerordentlich geringen Dichte der Störatome findet ein direkter Elektronenübergang zwischen den Störtermen im allgemeinen nicht statt. Das vom Donator abgelöste Elektron wird also an das Leitungsband abgegeben. Wird ein ionisiertes Störatom wieder neutralisiert, so geschieht das durch ein vom Leitungsband herunterfallendes Elektron. Deshalb sind die einzelnen Störniveaus nicht miteinander verbunden, sondern als lokalisierte Terme eingetragen.

Die Anregungsenergie eines neutralen Akzeptors der III. Gruppe ist ebenfalls nur etwa 0,01 eV, d. h. diese Elektronenzustände liegen, wie in Abb. 2.21f und 2.21g angedeutet, nur knapp über dem oberen Rand des Valenzbandes. $A^\times$ bedeutet einen unbesetzten, A^- einen mit einem Elektron besetzten Akzeptor. Dieses Elektron hat im Valenzband ein bewegliches Loch hinterlassen.

Sind Donatoren und Akzeptoren gleichzeitig und in gleicher Dichte anwesend, so rekombinieren die von den Donatoren im Leitfähigkeitsband geschaffenen Elektronen mit den von den Akzeptoren im Valenzband erzeugten Löchern (Abb. 2.22). Wie bereits erwähnt, bleiben in diesem Fall lediglich ionisierte Störatome, aber keine Leitfähigkeitsträger übrig. Gleichbedeutend damit ist auch die Aussage, daß die von den Donatoren gebildeten frei beweglichen Leitfähigkeitselektronen die noch neutralen Akzeptoren besetzen und damit einer Besetzung dieser Terme mit Elektronen aus dem Valenzband

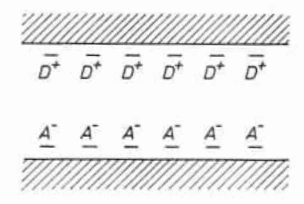


Abb. 2.22.
Bändermodell bei vollständiger Kompensation.

zuvorkommen oder, daß das vom Akzeptor geschaffene Loch zum neutralen Donator diffundiert und dort mit dem überzähligen Elektron rekombiniert. Wegen der geringen Störstoffdichten ist es unwahrscheinlich, daß das Elektron eines neutralen Donators ohne Zwischenstufen auf den neutralen Akzeptor übergeht.

Schließlich haben wir noch anzumerken, daß auch andere Stoffe außer denjenigen der III. und V. Gruppe, die wir normale Störstoffe nennen wollen, Störterme bilden, die zum Teil viel weiter von den Bandrändern entfernt sind, wie aus Abb. 2.23 ersichtlich ist.

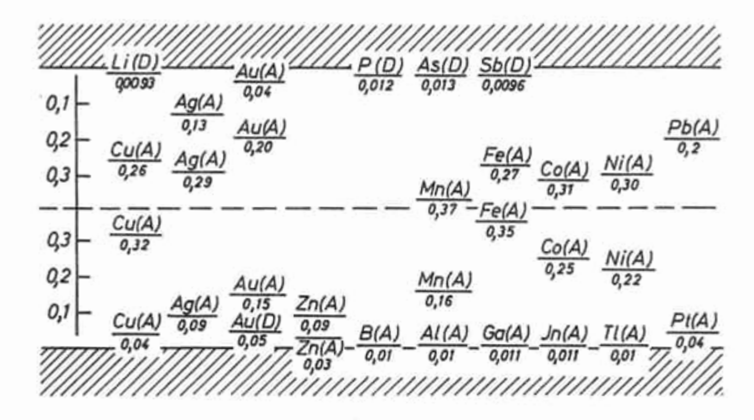


Abb. 2.23.
Störterme von Fremdstoffen in Germanium nach Conwell [21]. Die Zahlen bedeuten Energien in eV.

Die Einzelheiten des Einbaues dieser anomalen Störstoffe im Ge-Gitter gehen über den Rahmen einer Einführung hinaus; sie sollten lediglich der Vollständigkeit halber erwähnt werden.

Drittes Kapitel

Einführung in die wellenmechanischen Grundlagen der Halbleiterphysik

Das vorliegende Kapitel soll die oben ausgeführten Modellvorstellungen theoretisch mit Hilfe der Wellenmechanik unterbauen und somit das Verständnis für die Vorgänge im Halbleiter vertiefen. Zum Zwecke einer ersten Orientierung kann allerdings dieses Kapitel vorläufig übergangen werden, ohne daß das Verständnis der nachfolgenden Kapitel darunter leiden wird. Andererseits verbieten der beschränkte Umfang dieses Buches und die Kompliziertheit der Materie auch hier, mehr als eben nur eine Einführung zu geben. Eine eingehende Darstellung des Stoffes findet man z. B. bei E. SPENKE, „Elektronische Halbleiter" (Springer-Verlag).

3.1 Das freie Elektron

Bekanntlich wird die Bewegung eines Elektrons beschrieben durch eine Wellenfunktion, d. h. durch eine Funktion des Ortes und der Zeit $\Psi(x, y, z, t)$, die sogenannte SCHRÖDINGER-Funktion. Es ist $\Psi\Psi^* dV$ (mit Ψ^* gleich dem konjugiert komplexen Wert von Ψ) gleich der Wahrscheinlichkeit, das Elektron im Volumenelement dV aufzufinden. Dann dient

$$\int_{\text{Gesamtraum}} \Psi\Psi^* \, \mathrm{dV} = 1 \qquad \text{(Gl. 3.1)}$$

zur Normierung von Ψ. Diese Wellenfunktion muß die Differentialgleichung der Wellenmechanik, die SCHRÖDINGER-Gleichung, befriedigen:

$$-\frac{\hbar^2}{2m} \Delta\Psi + U\Psi = i\hbar \frac{\partial\Psi}{\partial t} \qquad \text{(Gl. 3.2)}$$

Hier ist $\hbar = h/2\pi$, h die PLANCKsche Konstante und m die Elektronenmasse. U ist die potentielle Energie, die im allgemeinen eine Funktion von Ort und Zeit $U(x, y, z, t)$ ist und die Form der Ψ-Funktion entscheidend bestimmt.

Für stationäre Fälle ist $U(x, y, z)$ unabhängig von der Zeit, wodurch Gl. 3.2 separierbar wird mit dem Produktansatz

$$\Psi = \psi(x, y, z) \cdot \vartheta(t) \quad \text{(Gl. 3.3)}$$

Nach Einsetzen von Gl. 3.3 in Gl. 3.2 und Division durch Ψ ergibt sich

$$i\hbar \frac{\partial \vartheta}{\partial t} \frac{1}{\vartheta} = E \quad \text{(Gl. 3.4)}$$

und

$$-\frac{\hbar^2}{2m} \frac{\Delta \psi}{\psi} + U = E$$

oder

$$\frac{\hbar^2}{2m} \Delta \psi + (E - U)\psi = 0 \quad \text{(Gl. 3.5)}$$

E ist ein Parameter, der weder vom Ort noch von der Zeit abhängt.

Die Lösung von Gl. 3.4 ist:

$$\vartheta \sim e^{-i\frac{E}{\hbar}t} \quad \text{(Gl. 3.6)}$$

Sie gibt einen zeitlich periodischen Vorgang wieder mit der Frequenz

$$\nu = \frac{1}{2\pi} \frac{E}{\hbar} = \frac{E}{h} \quad \text{(Gl. 3.7)}$$

Durch Vergleich mit der PLANCKschen Gleichung

$$E = h\nu \quad \text{(Gl. 3.8)}$$

erkennt man, daß unser Separationsparameter E die Bedeutung der Gesamtenergie des durch die Ψ-Funktion beschriebenen Elektrons hat.

Gl. 3.5 ist die zeitunabhängige SCHRÖDINGER-Gleichung, die wir zunächst für das Problem des freien Elektrons lösen wollen. Wir haben dabei ein räumlich konstantes Potential U anzusetzen, das auch Null sein kann. Dann stellt Gl. 3.5 die Differentialgleichung einer räumlich periodischen Funktion dar. Eine Lösung ist

$$\psi \sim e^{ikr} \quad \text{(Gl. 3.9)}$$

mit

$$k^2 = \frac{2mE}{\hbar^2}$$

oder

$$E = \frac{\hbar^2 k^2}{2m} \quad \text{(Gl. 3.10)}$$

Und mit Gl. 3.3 und 3.6 erhält man den Ausdruck für eine ebene Welle

$$\Psi = \psi\,\vartheta = \text{const} \cdot e^{i(\boldsymbol{k}\,\boldsymbol{r} - Et/\hbar)} \qquad \text{(Gl. 3.11)}$$

die sich in Richtung des Wellenvektors $\boldsymbol{k}$ mit der Wellenlänge $\lambda = 2\pi/k$ und der Frequenz von Gl. 3.7 ausbreitet. Ein Vergleich der Gl. 3.10 mit dem Ausdruck für die kinetische Energie in der klassischen Mechanik zeigt, daß der Impuls des freien Elektrons beschrieben wird durch

$$\boldsymbol{P} = \hbar\,\boldsymbol{k} \qquad \text{(Gl. 3.12)}$$

Dies ist nichts anderes als die DEBROGLIE-Bedingung $P = h/\lambda$.

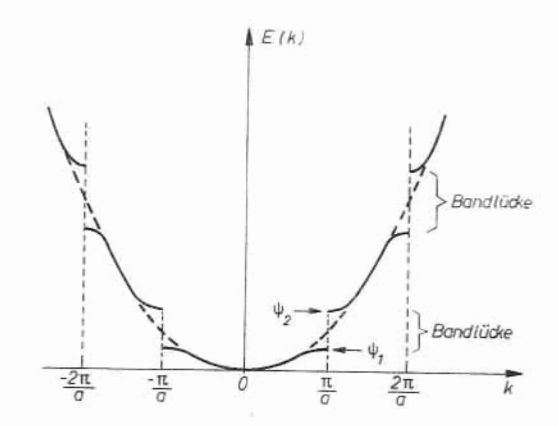


Abb. 3.1.
Energie als Funktion des Wellenvektors.
--- freie Elektronen, ——— Kristallelektronen.

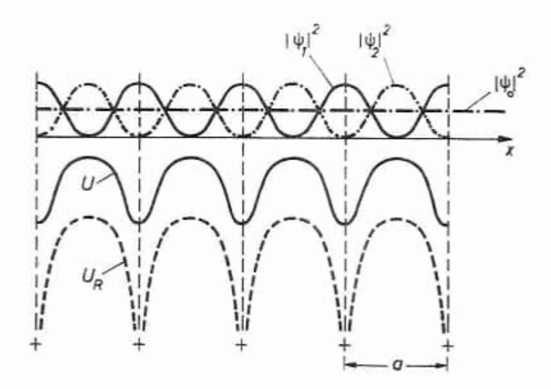


Abb. 3.2.
Potential und Elektronendichte im Kristallgitter.

Der parabolische Zusammenhang zwischen der Energie und dem Wellenvektor des freien Elektrons ist als gestrichelte Kurve in Abb. 3.1 dargestellt.

Nach dieser Einleitung wollen wir uns unserem eigentlichen Problem, nämlich dem des Kristallelektrons, zuwenden.

3.2 Das Kristallelektron in der Näherung schwacher Bindung

Das Problem des Kristallelektrons ist nicht geschlossen lösbar. Man ist deshalb auf verschiedene Näherungsverfahren angewiesen. Die eine dieser Näherungen, die BRILLOUINsche Näherung, geht aus von freien Elektronen, deren Funktion $E(k)$ in Abb. 3.1 gestrichelt dargestellt ist. Wird ein Strahl solcher Elektronen durch das Kristallgitter geschossen, so wird dessen Fortpflanzung durch die räumlich-periodischen Potentialänderungen beeinflußt, die durch die positiven Atomrümpfe hervorgerufen werden. Das Potential der positiven Rümpfe ist in Abb. 3.2 als U_R gestrichelt eingezeichnet. Seine Wirkung auf ein bestimmtes Elektron wird aber zum Teil abgeschirmt durch die Ladungen der anderen Elektronen. Es wird jedoch angenommen, daß diese Abschirmung an entsprechenden Punkten im Gitter etwa gleichgroß ist und somit nur eine Verschmierung (U in Abb. 3.2) der Rumpfpotentiale bewirkt in bezug auf das eine Auf-Elektron (Einelektronennäherung).

An diesen Potentialschwellen tritt eine Beugung der Elektronenwelle ein, die im allgemeinen gering ist, aber für bestimmte Wellenlängen und Richtungen zu BRAGG-Reflexion und damit durch Interferenz mit der einfallenden Welle zu stehenden Wellen führt. Wir wollen hier nur den eindimensionalen Fall, d. h. ein Liniengitter, betrachten und schreiben dafür die BRAGG-Beziehung für eine solche Interferenz:

$$n\lambda = 2a \qquad \text{(Gl. 3.13)}$$

wo a die Gitterkonstante ist, oder

$$k = \frac{2\pi}{\lambda} = n\,\frac{\pi}{a}; \qquad n = \pm 1, \pm 2, \pm 3, \ldots \qquad \text{(Gl. 3.14)}$$

Bei diesen Werten von k weicht das Verhalten des Kristallelektrons also stark von dem des freien Elektrons ab. Die Kurve der Abb. 3.1 erleidet bei diesen k-Werten Diskontinuitäten.

In Abb. 3.2 sind die zeitlich gemittelten Elektronendichten einer laufenden Welle $|\psi_0|^2$ und der beiden möglichen stehenden Wellen $|\psi_1|^2$ und $|\psi_2|^2$ für $k = \frac{\pi}{a}$ schematisch aufgezeichnet. ψ_1 und ψ_2 erhält man mit Hilfe der quantenmechanischen Störungsrechnung. Man sieht, daß $|\psi_1|^2$ seine Elektronendichten-Maxima an der Stelle der positiven Atomrümpfe hat, während $|\psi_2|^2$

dort seine Minima aufweist. Demzufolge ist die elektrostatische Energie des Schwingungszustandes ψ_1 tiefer als ψ_2. Wir haben also zwei diskrete Energiewerte für ein und dasselbe k, die sich beide von der Energie der laufenden Welle (freies Elektron) ψ_0 desselben Wellenvektors unterscheiden. In den k-Gebieten zwischen den Werten $\frac{n\pi}{a}$ verhalten sich die Elektronen annähernd wie freie Elektronen.

Dies führt, wie in Abb. 3.1 dargestellt ist, auf Bänder und Energielücken, wie wir sie schon in Kapitel 2.4.3 besprochen haben.

3.3 Das Kristallelektron in der Näherung starker Bindung

Die andere, sogenannte BLOCHsche Näherung ist ebenfalls eine Einelektronennäherung und geht von Elektronen aus, die stark an die Atome gebunden sind; d. h. man setzt hier als nullte Näherung die Eigenfunktionen ψ_{At} der isolierten Atome an. Dazu gehören bekanntlich bestimmte Energieeigenwerte E_{At}. Den Elektronen in den isolierten Atomen stehen also nur diskrete Energiezustände zur Verfügung.

Betrachten wir N gleichartige Atome, die zunächst noch weit voneinander entfernt sein sollen, so befinden sich deren Energieterme alle auf gleicher Energie, d. h. die Terme des Gesamtsystems sind N-fach entartet. Werden nun die Atome einander so nahe gebracht, daß sie durch Austausch ihrer Elektronen in Wechselwirkung treten können, dann hebt sich die Entartung auf, und die Energieterme spalten in Bänder auf (siehe Abb. 3.3). Diese Aufspaltung ist ganz analog zur Aufspaltung der Eigenfrequenzen gekoppelter mechanischer Pendel. Sie wird um so größer sein, je enger die Koppelung ist.

Somit liefern die BRILLOUINsche und die BLOCHsche Näherung qualitativ dasselbe Ergebnis. Um zu quantitativen Aussagen zu kommen, genügt aber nicht eines der beiden Verfahren allein. Vielmehr muß in komplizierten, von experimentellen Ergebnissen unterstützten Näherungsschritten von beiden Seiten an das Problem herangegangen werden. Dies soll hier nicht weiter verfolgt werden. Wir wollen vielmehr die oben erwähnte BLOCHsche Näherung etwas eingehender betrachten, um das prinzipielle Verhalten der Kristallelektronen an diesem Beispiel besser verstehen zu können.

Wir bauen also unsere ψ-Funktion an einer Stelle x in unserem linearen Gitter aus den atomaren Funktionen ψ_{At} aller benachbarten Atome auf.

$$\begin{aligned}\psi(x) = C_0\,\psi_{\mathrm{At}}(x) + C_1\,\psi_{\mathrm{At}}(x-a) + C_2\,\psi_{\mathrm{At}}(x-2a) + \dots\\ + C_{-1}\,\psi_{\mathrm{At}}(x+a) + C_{-2}\,\psi_{\mathrm{At}}(x+2a) + \dots\\ = \sum_{l=-\infty}^{l=+\infty} C_l\,\psi_{\mathrm{At}}(x-la) \qquad \text{(Gl. 3.15)}\end{aligned}$$

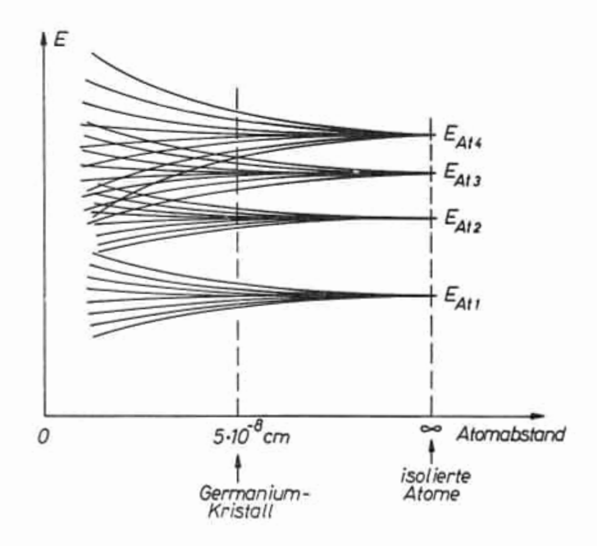


Abb. 3.3.
Aufspaltung der atomaren Terme in Bänder.

Die Koeffizienten C_l sollen nun mit Hilfe einer quantenmechanischen Störungsrechnung ermittelt werden.

Wir setzen Gl. 3.15 in die SCHRÖDINGER-Gleichung 3.5 ein:

$$\sum_l C_l \left\{ \frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + [E - U(x)] \right\} \psi_{At}(x - la) = 0 \qquad \text{(Gl. 3.16)}$$

Andererseits können wir auch die ψ_{At} einzeln in Gl. 3.5 einsetzen, wenn wir für U das Potential des isolierten Atoms U_{At} wählen

$$\left\{ \frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + [E_{At} - U_{At}(x - la)] \right\} \psi_{At}(x - la) = 0 \qquad \text{(Gl. 3.17)}$$

Wird nun Gl. 3.17 für alle Werte l angeschrieben, jeweils mit C_l multipliziert und die Summe aller dieser Gleichungen gebildet, so erhält man

$$\Sigma\, C_l \left\{ \frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + [E_{At} - U_{At}(x - la)] \right\} \psi_{At}(x - la) = 0 \qquad \text{(Gl. 3.18)}$$

Wir ziehen Gl. 3.18 von Gl. 3.16 ab, multiplizieren das Ergebnis mit dem konjugiert komplexen Wert einer der $\psi_{\mathrm{At}}(x - la)$, z. B. mit $\psi^*_{\mathrm{At}}(x - ra)$, und integrieren über den ganzen unendlich gedachten Kristall. Dann ergibt sich:

$$\sum_l C_l \left\{ \int_{-\infty}^{+\infty} \psi^*_{\mathrm{At}}(x - ra)\,[E - E_{\mathrm{At}}]\,\psi_{\mathrm{At}}(x - la)\,dx \right. \qquad \text{(Gl. 3.19)}$$

$$\left. - \int_{-\infty}^{+\infty} \psi^*_{\mathrm{At}}(x - ra)\,[U(x) - U_{\mathrm{At}}(x - la)]\,\psi_{\mathrm{At}}(x - la)\,dx \right\} = 0$$

Es wird nun in dieser Näherung im ersten Glied von Gl. 3.19 die Überlappung von zwei benachbarten ψ_{At}-Funktionen vernachlässigt, so daß ein Beitrag zu der Summe nur von dem Term mit $l = r$ entsteht, wobei für das Integral die Norminierungsbedingung Gl. 3.1. gilt. Beim zweiten Glied von Gl. 3.19 wird $[U - U_{\mathrm{At}}]$ für $x = la = ra$ sehr klein, so daß der Term $l = r$ nicht immer groß gegenüber allen anderen ist. Deshalb werden die Glieder $l = r - 1$ und $l = r + 1$ hier mitberücksichtigt. Man erhält dann

$$C_r[E - E_{\mathrm{At}}] - C_{r-1}A - C_rC - C_{r+1}A = 0 \qquad \text{(Gl. 3.20)}$$

Dabei wurden die folgenden Abkürzungen eingeführt:

$$A \equiv \int_{-\infty}^{+\infty} \psi^*_{\mathrm{At}}(x - ra)\,[U(x) - U_{\mathrm{At}}(x - (r-1)\mathrm{a})]\,\psi_{\mathrm{At}}(x - (r-1)\,a)\,dx \qquad \text{(Gl. 3.21)}$$

Diese Größe repräsentiert die durch Austausch von Elektronen zwischen benachbarten Atomen entstehende Bindungsenergie und wird Austauschintegral genannt.

$$C \equiv \int_{-\infty}^{+\infty} \psi^*_{\mathrm{At}}(x - ra)\,[U(x) - U_{\mathrm{At}}(x - ra)]\,\psi_{\mathrm{At}}(x - ra)\,dx \qquad \text{(Gl. 3.22)}$$

Dieses sogenannte COULOMBintegral stellt die COULOMBsche Energie des Elektrons im Felde des Störanteils $U - U_{\mathrm{At}}$ dar.

Die Gl. 3.20 stellt ein unendliches Gleichungssystem für die unendlich vielen C_l dar. Wir wollen die Lösung hier angeben:

$$C_l \sim e^{ikla} \qquad \text{(Gl. 3.23)}$$

k soll zunächst eine beliebige Konstante sein. Durch Einsetzen wird gezeigt, daß der Ausdruck Gl. 3.23 tatsächlich die Gl. 3.20 befriedigt. Außerdem erhält man dadurch einen Zusammenhang zwischen E und k.

$$E = E_{\mathrm{At}} + C + 2A\cos ka \qquad \text{(Gl. 3.24)}$$

In Abb. 3.4 ist diese Funktion aufgetragen, und zwar für zwei verschiedene atomare Eigenwerte E_{At_1} und E_{At_2} (und verschiedene C- und A-Werte). Man sieht, daß das letzte Glied, das „Austauschglied", die diskreten atomaren Eigenwerte in Bänder überführt, wie es am Eingang dieses Abschnittes schon angedeutet wurde. Die ganze Breite des Energiebandes wird in beiden Richtungen einmal überstrichen, wenn die Größe k die Werte von

$$-\frac{\pi}{a} \text{ bis } +\frac{\pi}{a}$$

durchläuft. Man sieht auch, daß die Breite der Bänder proportional zur Bindungsenergie A ist.

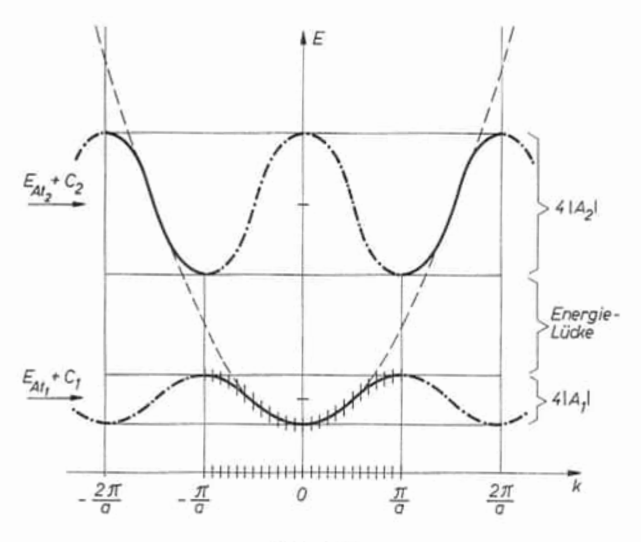


Abb. 3.4.
Energie als Funktion des Wellenvektors in der BLOCHschen Näherung.

Die Bedeutung von k wird klar, wenn wir die vollständige SCHRÖDINGER-Funktion anschreiben, indem wir Gl. 3.23, 3.15 und 3.6 in Gl. 3.3 einsetzen:

$$\Psi(x,t) = c \sum_l \psi_{\mathrm{At}}(x - la)\, e^{-ik(x-al)}\, e^{i(kx-Et/\hbar)} \qquad \text{(Gl. 3.25)}$$
$$= u(x,k)\, e^{i(kx-Et/\hbar)}$$

Die Amplitudenfunktion $u\,(x, k)$ ist periodisch in x mit der Periode a, weil die $\sum_l$ von $l = -\infty$ bis $l = +\infty$ geht. Wir haben also wieder eine Welle, die aber jetzt im Gegensatz zur freien Elektronenwelle eine räumlich gitterperiodisch

modulierte Amplitude besitzt; d. h. die Wahrscheinlichkeitsdichte ist an gewissen gitterperiodischen Stellen, z. B. an den Atomrümpfen, im Mittel höher als an anderen Stellen. Die Funktion $\Psi(x,t)$ der Gl. 3.25 wird als BLOCH-Funktion bezeichnet. Wir sehen nun auch, daß die Größe k die Bedeutung eines Wellenvektors hat.

Bei Übertragung dieser Gleichungen auf ein dreidimensionales Gitter geht Gl. 3.25 in den Ausdruck einer ebenen Welle über.

$$\Psi(\boldsymbol{r},t) = u(\boldsymbol{r},\boldsymbol{k})\, e^{i(\boldsymbol{k}\boldsymbol{r}-Et/\hbar)} \qquad \text{(Gl. 3.26)}$$

Nachdem nun die Bedeutung von k festliegt, zeigt ein Vergleich von Abb. 3.4 mit Abb. 3.1, daß die Ähnlichkeit der Ergebnisse beider Näherungen vollständig ist, wenn wir die strichpunktierten Teile der Cosinuslinien weglassen. Dies ist erlaubt, weil sich die E- und ψ-Werte des Gebietes $-\frac{\pi}{a} < k < +\frac{\pi}{a}$, der sogenannten ersten Zone, wiederholen, wenn k-Werte aus anderen Gebieten eingesetzt werden. Das ist leicht zu sehen, wenn $\psi(x)$ in der Form der Gl. 3.15 geschrieben wird.

$$\psi(x) \sim \sum_l \psi_{\mathrm{At}}(x - la)\, e^{ikla} \qquad \text{(Gl. 3.27)}$$

Betrachten wir nämlich einen k-Wert, der außerhalb der ersten Zone liegt, z. B. $k = +\frac{\pi}{a} + \varepsilon$, so ergibt das, in Gl. 3.27 eingesetzt, dieselbe Eigenfunktion wie $k = -\frac{\pi}{a} + \varepsilon$, was innerhalb der ersten Zone liegt. Es sind also in der sogenannten ersten BRILLOUIN-Zone,

$$-\frac{\pi}{a} \leq k \leq +\frac{\pi}{a} \qquad \text{(Gl. 3.28)}$$

schon alle sinnvollen k-Werte enthalten. Man kann auch diese Zone in das Gebiet $-\frac{2\pi}{a} \leq k \leq -\frac{\pi}{a}$ und $+\frac{\pi}{a} \leq k \leq +\frac{2\pi}{a}$ oder in eine höhere Zone zeichnen, wie das in Abb. 3.4 für das zweite Band geschehen ist. Gl. 3.27 zeigt auch, daß für $k = 0$ einerseits und für $k = \pm\frac{\pi}{a}$ andererseits stehende Wellen (vgl. ψ_1 und ψ_2 in Abb. 3.2.) auftreten.

Die Ähnlichkeit der beiden Näherungen ist natürlich nur dann gewährleistet, wenn das Vorzeichen von A in aufeinanderfolgenden Bändern abwechselt. Das braucht aber durchaus nicht immer so zu sein. Insbesondere werden die $E(\boldsymbol{k})$-Funktionen bei reellen dreidimensionalen Gittern sehr viel komplizierter, worauf wir weiter unten noch eingehen wollen.

3.4 Zyklische Randbedingung und Besetzung der Bänder

Beim Aufsuchen der ψ-Funktionen für das Kristallelektron haben wir bisher noch keine Randbedingung berücksichtigt. Wir wollen das jetzt nachholen, wobei sich zeigen wird, daß dadurch die Zahl der Eigenfunktionen ψ und damit auch die Zahl der Eigenwerte E und k beschränkt wird. Wir werden also sehen, daß die BRILLOUIN-Zone $-\frac{\pi}{a} \leq k \leq +\frac{\pi}{a}$ nicht kontinuierlich mit Quantenzuständen gefüllt ist, sondern nur eine beschränkte, aber sehr große Anzahl diskreter Eigenfunktionen, d. h. diskreter k-Werte, enthält.

Die Randbedingung unseres idealisierten Falles ist ganz einfach die, daß das Elektron sich in einem unendlichen Kristall befinden soll. Das wird mathematisch folgendermaßen formuliert: Man rechnet zwar mit einem endlich großen Kristall (der allerdings sehr viele Atome enthält), dem sogenannten Grundgebiet. An der Begrenzung dieses Grundgebietes sollen aber dann die ψ-Funktionen so verlaufen, als ob daran dasselbe Grundgebiet sich wieder anschließen würde. Im eindimensionalen Fall bedeutet dies folgendes: Das Grundgebiet enthalte N Atome im Abstand a, hat also eine Länge von

$$G = N a \qquad \text{(Gl. 3.29)}$$

Dann soll sein

$$\psi(0) = \psi(+G) \qquad \text{(Gl. 3.30)}$$

Die Wellenfunktionen sollen sich so verhalten, wie wenn unser eindimensionales Gitter im Kreis herum zusammengebogen sei und Anfang und Ende miteinander verknüpft seien. Wir sprechen deshalb von einer zyklischen Randbedingung. Diese Überlegung läßt sich gedanklich leicht (anschaulich etwas schwerer) auf das dreidimensionale Grundgebiet übertragen.

Wenden wir Gl. 3.30 auf Gl. 3.25 an, so sehen wir, daß $u(0) = u(+G)$ wegen der Periodizität von u und der Ganzzahligkeit von N erfüllt ist. Wir müssen also nur noch fordern

$$e^{ikG} = 1$$

Das führt auf die schon erwähnten diskreten Werte von k.

$$k = n\frac{2\pi}{G}; \qquad n = 0, \pm 1, \pm 2, \ldots \qquad \text{(Gl. 3.31)}$$

Diese k-Werte verteilen sich in gleichen Abständen $2\pi/G$ über eine BRILLOUIN-Zone, deren Breite $2\pi/a$ ist (Abb. 3.4). Somit ergibt sich die Zahl der möglichen k-Werte in der ganzen BRILLOUIN-Zone und damit auch die Zahl der Energiezustände im ganzen Band zu

$$\frac{2\pi/a}{2\pi/G} = \frac{Na}{a} = N$$

Nach dem Pauli-Prinzip kann jeder Energiezustand zwei Elektronen entgegengesetzten Spins aufnehmen. Es können somit in jedem Band $2N$, also doppelt so viele Elektronen Platz finden als Atome im Gitter sind. Ist bei einem Stoff z. B. ein Elektron pro Atom vorhanden, dann wird das unterste Band halb gefüllt sein, sind es zwei, dann wird es gerade eben ganz gefüllt sein. Bei mehreren Elektronen pro Atom werden die höheren Bänder der Reihe nach von unten nach oben gefüllt. Bei ungerader Zahl der Elektronen pro Atom wird das obere Band halb, bei gerader Zahl ganz gefüllt sein.

Dadurch kann nun der Unterschied in der Leitfähigkeit der verschiedenen festen Stoffe erklärt werden. Elektrische Leitfähigkeit bedeutet Beschleunigung, d. h. Energieaufnahme der Elektronen im elektrischen Feld. Das ist aber nur möglich, wenn leere Zustände im selben Band vorhanden sind, in welche die Elektronen bei der Energieaufnahme hochsteigen können. Somit ergibt sich Leitfähigkeit nur bei halbleeren Bändern. Volle Bänder tragen nicht zur Leitfähigkeit bei. Stoffe mit einem halbleeren obersten Band sind Metalle. Stoffe mit vollem obersten Band sind Isolatoren oder Halbleiter (Abb. 3.5).

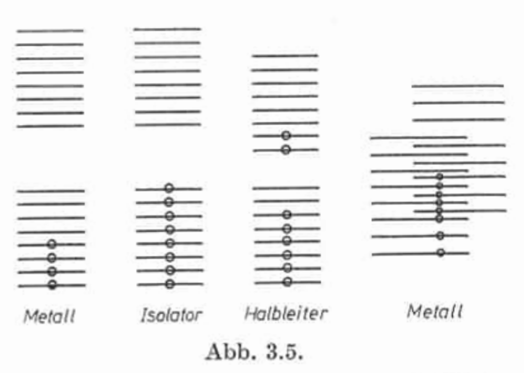


Abb. 3.5.
Besetzung der oberen Bänder bei Metallen, Halbleitern und Isolatoren.

Der Unterschied zwischen Halbleitern und Isolatoren ist eigentlich nur ein quantitativer: Bei Halbleitern ist die Energielücke über dem obersten besetzten Band, dem Valenzband, klein genug, so daß Valenzelektronen ins nächsthöhere leere Band, das Leitungsband, hochgehoben werden können. Diese Mechanismen sind im 2. Kapitel bereits ausreichend besprochen worden.

3.5 Bandstrukturen realer Gitter

Es treten nun allerdings in der Regel Komplikationen bezüglich der oben erwähnten einfachen Abzählungsvorschrift auf, und zwar dadurch, daß sich in realen dreidimensionalen Gittern die höheren Bänder vielfach überlappen (siehe auch Abb. 3.3). Tritt diese Überlappung z. B. zwischen einem vollen und einem leeren Band ein, so entsteht wieder ein halbvolles Gesamtband, das zu metallischen Eigenschaften führt (Abb. 3.5).

Wir interessieren uns hier aber insbesondere für die Bandstrukturen der Halbleiter, und speziell wiederum für unsere Prototypen Germanium und Silizium. Zur Einführung sei aber noch einmal ein Bänderpaar, wie es die Blochschen Näherung ergibt, in Abb. 3.6a betrachtet. Diese Abbildung

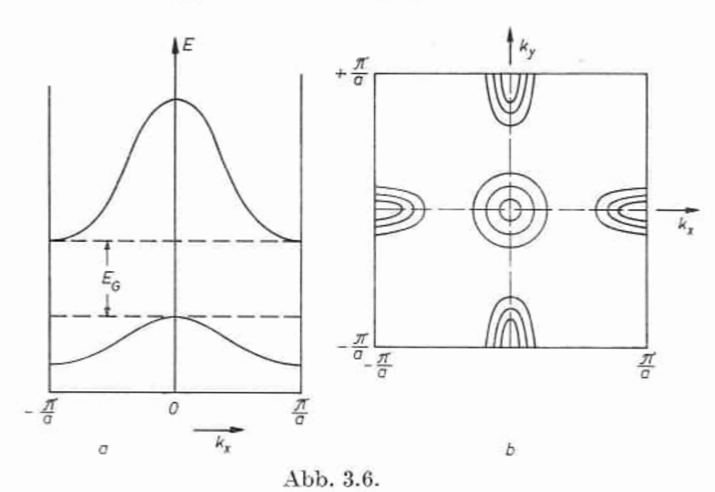


Abb. 3.6.
Die erste BRILLOUIN-Zone eines einfachen Gitters.

ist eine Wiederholung der Abb. 3.4, mit der Ausnahme, daß in beiden Bändern hier die Austauschenergie positiv sein soll. In Abb. 3.6b sind zwei Dimensionen der dreidimensionalen ersten BRILLOUIN-Zone aufgetragen mit den Linien gleicher Energie des oberen Randes des Valenzbandes (konzentrische Kreise in der Mitte) und des unteren Randes des Leitungsbandes (Ellipsen am Zonenrand); denn es sind ja die oberen bzw. unteren Bandränder, für die wir uns nach den Ausführungen von Abschnitt 2.4.3 beim Halbleiter besonders interessieren.

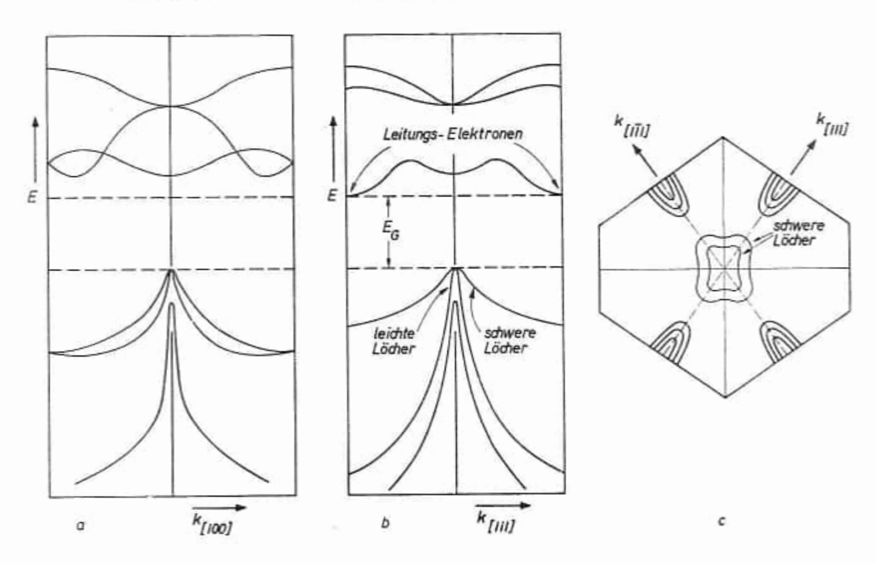


Abb. 3.7.
Die erste BRILLOUIN-Zone des Siliziums nach BROOKS [22].

Die gezeichneten Energielinien würden in einer dreidimensionalen Zeichnung als Energieflächen auftreten, im vorliegenden Fall in der Nähe des oberen Valenzbandrandes als konzentrische Kugeln, am unteren Leitungsbandrand als Rotationsellipsoide. Es ist klar, daß der Schnitt durch das Energiegebirge in Abb. 3.6a sich ändert, wenn anstatt entlang k_x entlang z. B. der Raumdiagonalen geschnitten wird.

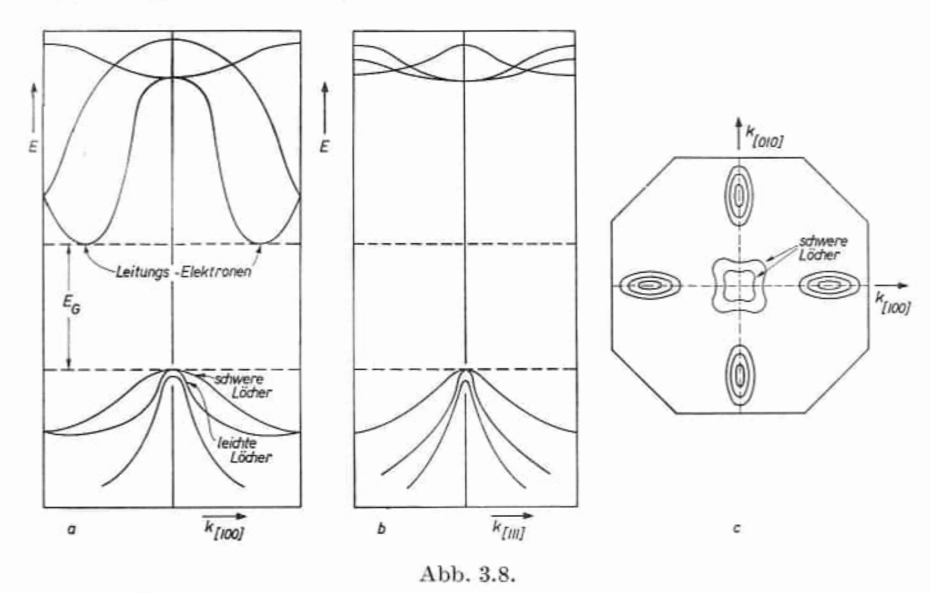


Abb. 3.8.
Die erste BRILLOUIN-Zone des Germaniums nach BROOKS [22].

In Abb. 3.7 und Abb. 3.8 sind nun die Verhältnisse bei Silizium und Germanium dargestellt. Man sieht, daß hier Valenzband und Leitungsband jeweils aus mehreren sich überlappenden Einzelbändern zusammengesetzt sind. Das Valenzband hat das Maximum bei $k = (0,0,0)$. Die Energieflächen sind aber nicht kugelförmig, sondern etwas einspringend geformt, was durch die Wechselwirkung der verschiedenen Teilbänder zustande kommt. Das Leitungsband hat bei beiden Stoffen mehrere Minima, wobei die tiefsten Minima, die den Bandabstand bestimmen, bei Silizium auf den Würfelachsen innerhalb der Zone liegen, bei Germanium auf den Raumdiagonalen und am Rande der Zone. Entsprechend sind auch für die zweidimensionalen Darstellungen der BRILLOUIN-Zonen bei Silizium die (001)-Ebene und bei Germanium die $(10\overline{1})$-Ebene gewählt worden. Man sieht auch, wie sich in der komplizierteren Form der BRILLOUIN-Zone die Kompliziertheit des Diamantgitters ausdrückt.

Selbstverständlich können die Funktionen $E(\boldsymbol{k})$ überall in der Nähe der Minima und Maxima durch quadratische Ausdrücke in $\boldsymbol{k}$ angenähert werden. Aus diesem Grunde verhalten sich die Elektronen und Löcher im Halbleiter wie quasi freie Teilchen (vgl. Gl. 3.10).

3.6 Das dynamische Verhalten des Kristallelektrons

Die durch die Gl. 3.25 oder 3.26 dargestellte BLOCHsche Funktion beschreibt eine unendlich ausgedehnte Welle, die zwar einen scharf bestimmten Wellenvektor k ($\sim$ Impuls), aber einen unendlich unbestimmten Ort x hat. Um die Unbestimmtheit δx des Ortes zu verkleinern, müssen wir den Impuls etwas unbestimmt machen, gemäß der HEISENBERGschen Unbestimmtheitsrelation:

$$\delta x \cdot \delta P = h \qquad \text{(Gl. 3.32)}$$

Dies bedeutet: Wir haben viele Wellen verschiedener Wellenlänge so zusammenzusetzen, daß durch Interferenz eine lokalisierte Aufenthaltswahrscheinlichkeit $\overline{\Psi}\overline{\Psi}^*$, ein sogenanntes Wellenpaket, herauskommt (Abb. 3.9).

Sind die Funktionen Ψ von Gl. 3.25 Lösungen der SCHRÖDINGER-Gleichung, so ist auch eine Summe dieser Funktionen eine Lösung. Wir bilden also das Wellenpaket als Summe aus vielen Wellen benachbarter Wellenlänge.

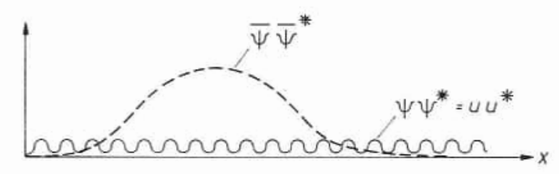


Abb. 3.9
Zum Aufbau eines Wellenpakets.

$$\overline{\Psi} = \frac{1}{2\,\delta k} \int_{k_0-\delta k}^{k_0+\delta k} \Psi(x,k)\, dk$$

$$= \frac{1}{2\,\delta k} \int_{k_0-\delta k}^{k_0+\delta k} u(x,k)\; e^{\,i(kx-Et/\hbar)}\, dk \qquad \text{(Gl. 3.33)}$$

Wir wollen nur eine kleine Umgebung von k_0 betrachten, also δk klein machen. Dann wird

$$u\,(x,k) \approx u\,(x,k_0) \qquad \text{(Gl. 3.34)}$$

und

$$E\,(k) \approx E\,(k_0) + \frac{\partial E}{\partial k}(k - k_0) \qquad \text{(Gl. 3.35)}$$

und damit

$$\overline{\Psi} = u\,(x,\,k_0)\; e^{\,i(k_0x-E(k_0)\,t/\hbar)} \cdot \frac{1}{2\cdot\delta k} \int_{-\delta k}^{+\delta k} e^{\,i\alpha\,(k-k_0)}\,\delta\,(k - k_0)$$

$$= u\,(x,\,k_0)\; e^{\,i(k_0x-E(k_0)\,t/\hbar)} \cdot \frac{\sin \alpha\cdot\delta k}{\alpha\cdot\delta k} \qquad \text{(Gl. 3.36)}$$

wobei

$$\alpha \equiv x - \frac{1}{\hbar} \frac{\partial E}{\partial k} t$$

Der Ort $\bar{x}$ des Wellenpakets ist dort, wo $\overline{\Psi}$ maximale Amplitude hat. Das ist nach Gl. 3.36 für $\alpha = 0$ erfüllt.

$$\bar{x} = \frac{1}{\hbar} \frac{\partial E}{\partial k} t \qquad \text{(Gl. 3.37)}$$

Daraus ergibt sich die Geschwindigkeit

$$v = \frac{\partial \bar{x}}{\partial t} = \frac{1}{\hbar} \frac{\partial E}{\partial k} \qquad \text{(Gl. 3.38)}$$

Diese Rechnung kann leicht auf drei Dimensionen ausgedehnt werden und ergibt

$$\boldsymbol{v} = \frac{1}{\hbar} \nabla_k E \qquad \text{(Gl. 3.39)}$$

Dabei ist

$$\nabla_k E = \frac{\partial E}{\partial k_x} \boldsymbol{e}_x + \frac{\partial E}{\partial k_y} \boldsymbol{e}_y + \frac{\partial E}{\partial k_z} \boldsymbol{e}_z$$

wo $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ die Einheitsvektoren sind. Man sieht unmittelbar aus Gl. 3.38 und Gl. 3.39, daß die Elektronen an den Bandrändern, d. h. in Maxima und Minima der Funktion $E(k)$ die Geschwindigkeit Null besitzen (stehende Wellen!).

Als Impuls ergibt sich dann

$$\boldsymbol{P} = m\,\boldsymbol{v} = \frac{m}{\hbar} \nabla_k E \qquad \text{(Gl. 3.40)}$$

Wird in diese Gleichung der Ausdruck der Energie für das freie Elektron (Gl. 3.10) eingesetzt, so erhalten wir Gl. 3.12, die aber nur für freie Elektronen gilt.

Wird das Elektron einer Kraft $\boldsymbol{K}$ unterworfen, z. B. durch ein elektrisches Feld, so nimmt seine Energie in der Zeit dt zu, und zwar um

$$dE = \boldsymbol{K}\,\boldsymbol{v}\,dt = \boldsymbol{K} \frac{1}{\hbar} \nabla_k E\,dt \qquad \text{(Gl. 3.41)}$$

Diese Änderung dE kann auch durch die zugehörige Änderung des Wellenvektors $\boldsymbol{k}$ ausgedrückt werden:

$$dE = \nabla_k E \frac{d\boldsymbol{k}}{dt} dt \qquad \text{(Gl. 3.42)}$$

Ein Vergleich beider Gleichungen ergibt

$$\boldsymbol{K} = \hbar \dot{\boldsymbol{k}} \qquad \text{(Gl. 3.43)}$$

Dies ist das Äquivalent der NEWTONschen Gleichung aus der klassischen Mechanik und entspricht ihr auch mit Gl. 3.12 für das freie Elektron. Ganz allgemein können wir die Größe $\hbar \boldsymbol{k}$ als „Kristallimpuls" bezeichnen. Dieser Kristallimpuls ist aber für das Kristallelektron nicht gleich dem Teilchenimpuls $\boldsymbol{P}$, der durch Gl. 3.40 gegeben ist.

Es läßt sich jedoch eine sogenannte „effektive Masse" definieren, so daß der Kristallimpuls gleich dem Produkt dieser Masse mit der Teilchengeschwindigkeit ist:

$$K = \hbar \dot{\boldsymbol{k}} \equiv m_{\text{eff}} \dot{\boldsymbol{v}} \qquad \text{(Gl. 3.44)}$$

Aus Gl. 3.38 erhalten wir für den eindimensionalen Fall

$$\dot{v} = \frac{\partial v}{\partial k} \dot{k} = \frac{1}{\hbar} \frac{\partial^2 E}{\partial k^2} \dot{k} \qquad \text{(Gl. 3.45)}$$

und durch Einsetzen in Gl. 3.44 erhält man

$$\frac{1}{m_{\text{eff}}} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k^2} \qquad \text{(Gl. 3.46)}$$

Die effektive Masse ist also umgekehrt proportional zur Krümmung der Energiefläche im k-Raum.

Am Rand eines Bandes kann E nach k entwickelt werden.

$$E = E_0 + \frac{\partial E}{\partial k}(k - k_0) + \frac{1}{2} \frac{\partial^2 E}{\partial k^2}(k - k_0)^2 + \dots$$

Da die erste Ableitung am Bandrand verschwindet, erhält man bei Vernachlässigung der höheren Glieder

$$E - E_0 = \frac{\hbar^2 (k - k_0)^2}{2 m_{\text{eff}}} \qquad \text{(Gl. 3.47)}$$

Aus dieser Gleichung wird durch Vergleich mit Gl. 3.10 die Bedeutung der effektiven Masse und des Kristallimpulses noch einmal klar.

Wir können Gl. 3.46 auf drei Dimensionen erweitern, wenn wir von Gl. 3.39 ausgehen und deren drei Komponenten nach der Zeit differenzieren.

$$\dot{v}_x = \frac{1}{\hbar}\left(\frac{\partial^2 E}{\partial k_x^2}\dot{k}_x + \frac{\partial^2 E}{\partial k_x \partial k_y}\dot{k}_y + \frac{\partial^2 E}{\partial k_x \partial k_z}\dot{k}_z\right)$$
$$\dot{v}_y = \frac{1}{\hbar}\left(\frac{\partial^2 E}{\partial k_y \partial k_x}\dot{k}_x + \frac{\partial^2 E}{\partial k_y^2}\dot{k}_y + \frac{\partial^2 E}{\partial k_y \partial k_z}\dot{k}_z\right)$$
$$\dot{v}_z = \frac{1}{\hbar}\left(\frac{\partial^2 E}{\partial k_z \partial k_x}\dot{k}_x + \frac{\partial^2 E}{\partial k_z \partial k_y}\dot{k}_y + \frac{\partial^2 E}{\partial k_z^2}\dot{k}_z\right)$$

oder in Tensorschreibweise:

$$\dot{v} = \frac{1}{\hbar}(T)\,\dot{k} \tag{Gl. 3.48}$$

Mit Gl. 3.44 erhält man entsprechend

$$\frac{1}{m_{\text{eff}}} = \frac{1}{\hbar^2}(T) \tag{Gl. 3.49}$$

wobei der Tensor (T) die Form hat

$$(T) \equiv \begin{pmatrix} \dfrac{\partial^2 E}{\partial k_x^2} & \dfrac{\partial^2 E}{\partial k_x \partial k_y} & \dfrac{\partial^2 E}{\partial k_x \partial k_z} \\ \dfrac{\partial^2 E}{\partial k_y \partial k_x} & \dfrac{\partial^2 E}{\partial k_y^2} & \dfrac{\partial^2 E}{\partial k_y \partial k_z} \\ \dfrac{\partial^2 E}{\partial k_z \partial k_x} & \dfrac{\partial^2 E}{\partial k_z \partial k_y} & \dfrac{\partial^2 E}{\partial k_z^2} \end{pmatrix} \tag{Gl. 3.50}$$

Durch entsprechende Wahl des Koordinatensystems kann erreicht werden, daß die Nicht-Diagonalglieder dieses Tensors verschwinden. Man sieht dann sehr deutlich, daß im Allgemeinfall für die drei Bewegungsrichtungen drei verschiedene effektive Massen auftreten

$$\begin{aligned} \dot{v}_x &= \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k_x^2} K_x = \frac{1}{m_x} K_x \\ \dot{v}_y &= \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k_y^2} K_y = \frac{1}{m_y} K_y \\ \dot{v}_z &= \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k_z^2} K_z = \frac{1}{m_z} K_z \end{aligned} \tag{Gl. 3.51}$$

Für eine beliebige Kraftrichtung wird deshalb die Beschleunigung des Kristallelektrons nicht immer in dieser Richtung erfolgen. Dies erscheint paradox. Die Ursache liegt aber darin, daß das Kristallelektron in erster Linie den Gitterkräften unterliegt und nicht wirklich frei ist, der von außen angelegten Kraft zu folgen. Als Vergleich aus der klassischen Mechanik sei ein schienengebundener Wagen erwähnt, der einer schräg zu den Schienen wirkenden Kraft nur dadurch zu folgen vermag, daß er sich in Richtung der Schienen bewegt. Außer der angelegten Kraft wirken noch die Zwangskräfte der Schienen.

Die Krümmung $\partial^2 E/\partial k^2$ ist um so stärker, die effektive Masse also um so kleiner, je größer die Austauschenergie A ist. Höhere Austauschenergie bedeutet aber schwächer gebundene Elektronen und breiteres Band (siehe Gl. 3.24, vgl. Abb. 3.4). Dies ist plausibel: Bei starker Bindung an die Atome (schmales Band) sind die Elektronen schwerer zu bewegen; sie sind träger (vgl. auch Gl. 3.55).

Wir haben in Abschnitt 3.5 gezeigt, daß die Energiefläche am unteren Rand des Leitungsbandes bei Ge und Si ein Rotationsellipsoid ist. Dem entspricht eine effektive Elektronenmasse m_l in der Längsrichtung des Ellipsoides, die größer ist als die (in allen Querrichtungen gleiche) transversale Elektronenmasse m_t. Diese Massen sind durch Zyklotronresonanzmessung bestimmt worden und in Tabelle 2 als Verhältnis zur Masse m des freien Elektrons aufgeführt.

Tabelle 2
Effektive Massen von Ge und Si

	Elektronen		Löcher	
	m_l/m	m_t/m	m_1/m	m_2/m
Ge	1,6	0,082	0,3	0,04
Si	0,97	0,19	0,5	0,16

Das merkwürdigste Ergebnis der obigen Gleichungen ist jedoch, daß für die Elektronen in den Zuständen am oberen Rand eines Bandes negative effektive Massen herauskommen (vgl. Gl. 3.46 und Abb. 3.6, 3.7 und 3.8). Auch hier müssen wir uns vor Augen halten, daß es eben im wesentlichen die Gitterkräfte sind, die durch BRAGGsche Interferenzen dafür sorgen, daß die Wellenpakete entgegen der äußeren Zusatzkraft laufen.

3.7 Elektronen- und Löcherleitung

In den Abschnitten 3.2 – 3.6 haben wir mit vollkommenen Kristallgittern gerechnet. Die dafür gewonnenen Elektronenwellen sind ungedämpft; die Wellenfunktionen enthalten kein Dämpfungsglied, und deshalb erfahren die daraus zusammengesetzten Wellenpakete, die Elektronen, im idealen Gitter keine Reibung. Eine Reibung und damit ein elektrischer Widerstand kommt erst durch den Einfluß der Unvollkommenheit des Gitters zustande, d. h. insbesondere durch die Verzerrung des Gitters infolge der thermischen Schwingungen der Gitteratome, durch eingebaute Verunreinigungen und durch strukturelle Fehlordnungen.

Wir wollen das Problem der Streuung hier nicht quantenmechanisch behandeln, sondern nur sehr vereinfacht im Teilchenbild andeuten. Man kann die Wirkung der Gitterunvollkommenheit auf die Elektronen beschreiben als Zusammenstöße, die die Elektronen mit den Verunreinigungsatomen und mit den Quanten der Gitterschwingungen, den Phononen, erleiden. Eine solche Elektronenbewegung ohne angelegtes Feld ist in Abb. 3.10a angedeutet. Wird nun ein elektrisches Feld angelegt, so überlagert sich die vom Feld hervorgerufene Geschwindigkeitskomponente $\boldsymbol{v}$ der thermischen Geschwindigkeit $\boldsymbol{v}_{th}$ (von Abb. 3.10a) zur Geschwindigkeit

$$\boldsymbol{v}_g = \boldsymbol{v}_{th} + \boldsymbol{v} \qquad \text{(Gl. 3.52)}$$

die in Abb. 3.10b gezeichnet ist. In der Zeit zwischen den Stößen wirkt nur die Beschleunigung durch das Feld gemäß Gl. 3.44.

$$\dot{\boldsymbol{v}} = -\frac{e\boldsymbol{F}}{m_{\text{eff}}} \qquad \text{(Gl. 3.53)}$$



Abb. 3.10.
Die Bahn eines Elektrons im realen Gitter, a) ohne Feld, b) mit Feld.

Wir nehmen nun an, daß das Elektron bei jedem Stoß so abgebremst oder beschleunigt wird, daß es im Mittel der thermischen Geschwindigkeitsverteilung $\boldsymbol{v}_{th}$ entspricht, d. h., daß es bei jedem Stoß die Existenz des Feldes $\boldsymbol{F}$ vergißt. Dann ist zur Zeit des Stoßes ($t = 0$) $\boldsymbol{v} = 0$. Die Zeit zwischen zwei Stößen sei im Mittel τ_s (ungeschickterweise „Stoßzeit" genannt). τ_s ist bei Zimmertemperatur etwa 10^{-12} sec. Dann ist zur Zeit $t = \tau_s$

$$\boldsymbol{v}(\tau_s) = \int_0^{\tau_s} \dot{\boldsymbol{v}}\, dt = -\frac{e\tau_s}{m_{\text{eff}}}\boldsymbol{F}$$

Der zeitliche Mittelwert von $\boldsymbol{v}$, d. h. der durch das Feld verursachten Zusatzkomponente der Geschwindigkeit, wäre dann die Hälfte von $\boldsymbol{v}(\tau_s)$. Eine genauere Rechnung, die die wirkliche Geschwindigkeitsverteilung berücksichtigt, gibt jedoch für diesen Mittelwert keinen Faktor 1/2, sondern

$$\overline{\boldsymbol{v}} = -\frac{e\tau_s}{m_{\text{eff}}}\boldsymbol{F} \qquad \text{(Gl. 3.54)}$$

Diese Gleichung ist identisch mit Gl. 2.10, wenn wir für $q = -e$ (Elektronen) und für die Beweglichkeit

$$\mu = \frac{e\tau_s}{m_{\text{eff}}} \qquad \text{(Gl. 3.55)}$$

setzen.

Es soll nun an Hand der einfachen Bandstruktur von Abb. 3.6 untersucht werden, wie sich die Gesamtheit der Elektronen eines Bandes unter dem Einfluß eines elektrischen Feldes verhält. Wir wollen zunächst ein fast leeres Band betrachten, wie es als Leitungsband in einem n-Leiter vorliegt, und wie es in Abb. 3.11a für den Fall $F = 0$ gezeichnet ist. Es sind nur die alleruntersten Zustände bis zu einer gewissen Linie besetzt. Die besetzten Zustände sind durch vertikale Striche markiert. Aus Gl. 3.38 kann für jeden besetzten Zustand die Geschwindigkeit v der Elektronen ermittelt werden. Für die linken Zustände

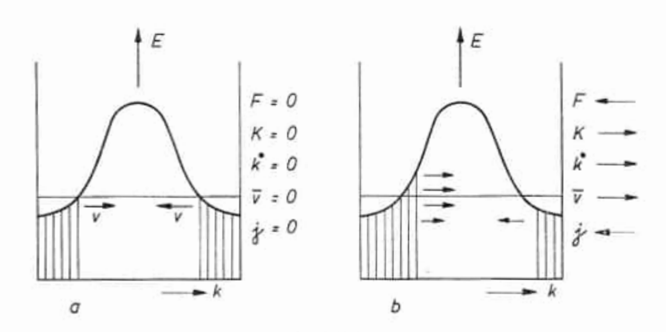


Abb. 3.11.
Elektronenverteilung in einem fast leeren Band.

$k < 0$ ist $\partial E/\partial k > 0$ also $v > 0$. Für die rechten Zustände $k > 0$ ist es umgekehrt. Wegen der Symmetrie von Abb. 3.11a heben sich beide Anteile auf, und man erhält $\bar{v} = 0$ und Stromdichte $j = 0$.

Wird nun ein von rechts nach links zeigendes Feld F angelegt, so wirkt auf die Elektronen eine Kraft $K = -eF$, die eine Verschiebung der k-Verteilung hervorruft. Gemäß Gl. 3.43 ergibt sich nämlich beim Einschalten ein gewisses $\dot{k}$, das dann aber durch Stöße aufgehoben wird und zu einer stationären Verschiebung der k-Besetzung führt, wie in Abb. 3.11b gezeichnet. Ebenfalls angedeutet sind die Geschwindigkeiten der verschiedenen Zustände, die sich jetzt bei der verschobenen Verteilung nur zum Teil aufheben. Wir erhalten ein $\bar{v}$ von links nach rechts und ein j in der Gegenrichtung, d. h. in der Richtung von F.

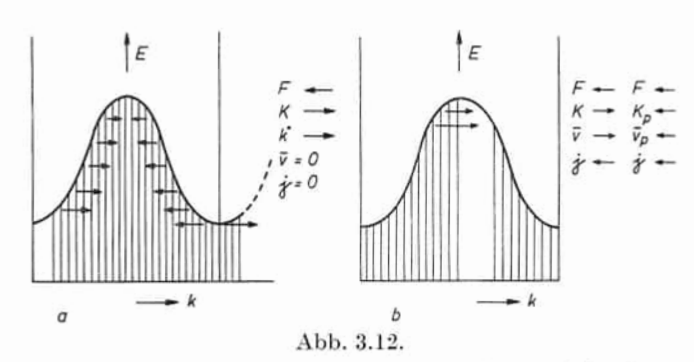


Abb. 3.12.
Elektronenverteilung in einem vollen Band (a) und in einem fast vollen Band (b).

Abb. 3.12a zeigt nun eine gleichartige Verschiebung in einem vollen Band. Die Zustände sind zum Teil nach rechts aus der BRILLOUIN-Zone „herausgelaufen". Wir wissen aber, daß nach Abschnitt 3.3 diese Zustände äquivalent denjenigen auf der linken inneren Seite der Zone sind; d. h. dieselben Zustände

laufen auf der linken Seite wieder herein. Dies oder die Bilanz der v-Werte in Abb. 3.12a zeigt, daß $\overline{v} = 0$, $j = 0$, d. h., daß sich im vollen Band Elektronen positiver und negativer Masse [positiver und negativer Krümmung der Funktion $E(k)$] gerade aufheben.

Schließlich seien an Hand von Abb. 3.12b die Verhältnisse in einem fast vollen Band betrachtet. Die Überlegung gilt für die Elektronen wie oben. Quantitativ läßt sich die Stromdichte aus derjenigen des ganzen, vollen Bandes j_v weniger der der fehlenden Elektronen anschreiben, wobei $j_v = 0$ ist, wie oben ausgeführt wurde, und wobei Gl. 3.54 und Gl. 3.55 berücksichtigt sind.

$$j = j_v - e\,p\,\frac{e\,\tau_s}{m_n}\,F = -e\,p\,\frac{e\,\tau_s}{m_n}\,F \qquad \text{(Gl. 3.56)}$$

Hier ist p die Zahl der unbesetzten Plätze im Band pro cm³ und m_n die effektive Masse, die die Elektronen in diesen Plätzen haben würden. Gemäß Gl. 3.46 ist m_n negativ. An dieser Stelle läßt sich zwanglos der Begriff des Defektelektrons oder Loches einführen. Man schreibt den leeren Plätzen positive Ladung $(+e)$ und eine effektive Masse m_p zu, die dem Betrag nach gleich m_n, aber positiv ist.

$$m_p = -m_n = +|m_n| \qquad \text{(Gl. 3.57)}$$

Die Änderung des Ladungsvorzeichens ändert Gl. 3.56 nicht; Einsetzen von Gl. 3.57 in 3.56 ergibt aber

$$j = +e\,p\,\frac{e\,\tau_s}{m_p}\,F = e\,p\,\mu_p\,F \qquad \text{(Gl. 3.58)}$$

Die durch dieses Modell entstehende Rechenvereinfachung wird deutlich an der zweiten Spalte von Abb. 3.12b, wo das Feld F, die Kraft auf das positiv geladene Loch K_p, die Geschwindigkeit der Löcher v_p und die Stromdichte alle in dieselbe Richtung zeigen.

In den Bänderstrukturen für Germanium und Silizium, die in Abb. 3.7 und 3.8 gezeigt sind, berühren sich im Maximum des Valenzbandes zwei Teilbänder. Eine Folge dieser Entartung ist eine leichte Einwölbung der Energiekugel im k-Raum, wie sie in den Abb. 3.7c und 3.8c etwas übertrieben gezeichnet ist. Das hat eine etwas anisotrope Löchermasse zur Folge. Wichtiger ist aber, daß die beiden Teilbänder bei $k = 0$ verschiedene Krümmungen haben. Infolgedessen gibt es in Ge und Si zweierlei Löcher, schwere und leichte, deren Massen als m_1 und m_2 in Tabelle 2 angegeben sind. Der Anteil der leichten Löcher an der Gesamtzahl ist bei Ge nur 5%, bei Si 15%. Da die effektiven Massen in die Beweglichkeit eingehen, und da die obigen Prozentsätze sich mit der Temperatur ändern, wirkt sich diese Tatsache insbesondere als Komplikation bei der Deutung der Hall-Effekt-Kurven (z. B. Abb. 2.15) aus. Für die Vorgänge in Gleichrichtern und Transistoren können wir jedoch in der Regel mit einer einheitlichen Beweglichkeit rechnen (siehe hierzu Rittner [23]).

Viertes Kapitel

Die Trägerkonzentrationen bestimmen die Eigenschaften des Halbleiters

4.1 Gleichgewicht

4.1.1 Zusammenhänge zwischen Elektronen-, Löcher- und Störstellendichten (Massenwirkungsgesetz). Im Kapitel 2, insbesondere Abschnitt 2.4.2.3, ist schon ein Zusammenhang zwischen den Trägerdichten und den Störstellendichten angedeutet worden. Dies soll nun in den folgenden Abschnitten exakter und allgemeiner behandelt werden. Unser Problem ist also, bei gegebener Dichte von Donatoren und Akzeptoren und bei gegebener Temperatur die Elektronen- und Löcherdichte zu berechnen.

Nach den Erörterungen in Kapitel 2 gibt es im Halbleiter grundsätzlich zwei verschiedene Prozesse, die zu einer elektrischen Leitfähigkeit führen: Eigenleitfähigkeit und Störleitfähigkeit. Im Fall der Eigenleitfähigkeit hatten wir in Abschnitt 2.4.1 festgestellt, daß die Generation von Elektron und Loch eine temperaturabhängige Konstante und die Rekombination proportional dem Produkt von Löcher- und Elektronendichte ist. Dies führt zu der Beziehung

$$pn = F_1(T) \qquad \text{(Gl. 4.1)}$$

Die Reaktion

$$\oplus + \ominus \rightleftarrows \text{gebundenes Valenzelektron} \qquad \text{(Gl. 4.2)}$$

führt auf Grund des Massenwirkungsgesetzes zu derselben Gleichung. Es ist nämlich:

$$\frac{p \cdot n}{\text{Dichte der Valenzelektronen}} = C(T) \qquad \text{(Gl. 4.3)}$$

Die Valenzelektronendichte ist von der Größenordnung der Dichte der Halbleiteratome, also etwa 10^{23} cm^{-3}. Sie kann durch die geringe Zahl der Löcher (10^{14} bis 10^{18} cm^{-3}) nicht wesentlich verändert werden und ist als Konstante zu betrachten. Somit ist Gl. 4.1 äquivalent mit Gl. 4.3 und als solche eine Folge der Reaktion Gl. 4.2. Da auch neben den Störleitungsreaktionen

$$D^{\times} \rightleftarrows D^{+} + \ominus \qquad \text{(Gl. 4.4)}$$

$$A^{\times} \rightleftarrows A^{-} + \oplus \qquad \text{(Gl. 4.5)}$$

die Reaktion Gl. 4.2 bestehen bleibt, gilt Gl. 4.1 auch bei Störleitung. Bei Eigenleitung ist $n = p = n_i$, deshalb gilt wegen Gl. 4.1 allgemein

$$p \cdot n = n_i^2 \qquad \text{(Gl. 4.6)}$$

Der Eigenleitungsprozeß sorgt also dafür, daß auch bei überwiegender Störleitung beide Arten von Ladungsträgern zugegen sind. Die Eigenleitungsdichte hängt vom Bandabstand E_G und von der Temperatur ab (vgl. Gl. 2.18).

Es sei $N_{D^\times}$ die Dichte der neutralen, N_{D^+} die der ionisierten Donatoren und entsprechend $N_{A^\times}$ und N_{A^-} für die Akzeptoren. Dann ist

$$N_{D^\times} + N_{D^+} = N_D \qquad \text{(Gl. 4.7)}$$

und

$$N_{A^\times} + N_{A^-} = N_A \qquad \text{(Gl. 4.8)}$$

Wenden wir auf die Reaktionen Gl. 4.4 und Gl. 4.5 das Massenwirkungsgesetz an, so erhalten wir

$$\frac{N_{D^\times}}{N_{D^+} \cdot n} = F_2(T) \qquad \text{(Gl. 4.9)}$$

und

$$\frac{N_{A^\times}}{N_{A^-} \cdot p} = F_3(T) \qquad \text{(Gl. 4.10)}$$

Schließlich gilt noch elektrische Neutralität

$$N_{D^+} + p = N_{A^-} + n \qquad \text{(Gl. 4.11)}$$

Aus den Gleichungen 4.7, 4.8, 4.9, 4.10 und 4.11 lassen sich $N_{D^\times}$, N_{D^+}, $N_{A^\times}$ und N_{A^-} eliminieren und man erhält

$$n - \frac{N_D}{1 + n F_2} = p - \frac{N_A}{1 + p F_3} \qquad \text{(Gl. 4.12)}$$

Aus dieser Gleichung kann zusammen mit Gl. 4.6 die Dichte der Elektronen und Löcher für alle Temperaturen und Störstellendichten N_D und N_A berechnet werden, falls $F_2(T)$ und $F_3(T)$ bekannt sind.

So ergibt z. B. $N_D = 0$, $N_A = 0$ mit Gl. 4.12 $p = n = n_i$ den Fall der reinen Eigenleitung. Bei reinem n-Leiter haben wir $N_A = 0$ und $n_i \ll n$ in die Gleichungen 4.12 und 4.6 einzusetzen und erhalten

$$n^2 = (N_D - n) \frac{1}{F_2} \qquad \text{(Gl. 4.13)}$$

was Gl. 2.19 in Abschnitt 2.4.2.3 entspricht, und woraus dort das Reserve- und Erschöpfungsverhalten des n-Halbleiters abgeleitet worden ist.

Ein weiterer wichtiger Fall ist der der gemischten Stör- und Eigenleitung des reinen n-Leiters ($N_A = 0$) bei hoher Temperatur, d. h. bei Erschöpfung. Dann ist $N_{D^+} \approx N_D$. Das ergibt aus Gl. 4.11

$$n = N_D + p \qquad \text{(Gl. 4.14)}$$

Für die Löcherdichte erhält man mit Gl. 4.6

$$p = \sqrt{\frac{N_D^2}{4} + n_i^2} - \frac{N_D}{2} \qquad \text{(Gl. 4.15)}$$

und näherungsweise für überwiegende Störleitung durch Entwickeln der Wurzel

$$p \approx \frac{n_i^2}{N_D} \qquad \text{(Gl. 4.16)}$$

Falls die Störleitung die Eigenleitung um einige Dekaden überwiegt, liegt die Dichte der kontrapolaren Leitfähigkeitsträger, beispielsweise bei Donatoren die der Defektelektronen, wir nennen sie kurz Minoritäten, um etwa denselben Faktor unter der Eigenleitungsdichte. Bei Leitfähigkeit und HALL-Effekt sind die Minoritäten zu vernachlässigen. Wir können sie aber hier nicht übergehen, weil sie bei dem Gleichrichter- und Transistoreffekt eine ganz entscheidende Rolle spielen.

Die eben geschilderten Verhältnisse liegen ähnlich wie bei einem Elektrolyten. So hat beispielsweise neutrales Wasser durch eine geringe Eigendissoziation nach der Reaktionsgleichung

$$H_2O \rightleftarrows H^+ + OH^-$$

eine geringe elektrische Leitfähigkeit. Der Eigendissoziationsprozeß läuft auch ab, wenn z. B. eine Säure (also Wasserstoffionen) zu neutralem Wasser gegeben wird. Das Produkt aus H^+ und OH^--Ionenkonzentration muß nach wie vor konstant bleiben, die OH^--Ionenkonzentration also abnehmen.

4.1.2 FERMI- und BOLTZMANN-Statistik. Im letzten Abschnitt haben wir lediglich durch Anwendung des Massenwirkungsgesetzes und der Neutralitätsbedingung einen sehr weitgehenden qualitativen und quantitativen Einblick in die Zusammenhänge zwischen den verschiedenen Konzentrationen erhalten. Offengeblieben sind noch die genauen temperaturabhängigen Massenwirkungskonstanten F_1, F_2 und F_3. Prinzipiell können auch diese Funktionen weitgehend auf thermodynamischem Wege berechnet werden. Es ist jedoch einfacher und aufschlußreicher, dies mit Hilfe der Statistik zu tun.

Im Abschnitt 2.4 haben wir schon mehrfach die BOLTZMANN-Statistik herangezogen, z. B. in Gl. 2.18. Strenggenommen gilt diese jedoch nicht, da sie z. B. für $T = 0$ alle Elektronen auf dem tiefsten Energieniveau versammeln würde, was auf Grund des PAULI-Prinzips unmöglich ist. Wird dieses Prinzip, wonach jeder Zustand nur mit zwei Elektronen entgegengesetzten Spins besetzt werden kann, und das quantenmechanische Prinzip der Ununterscheidbarkeit der Elektronen berücksichtigt, so kommt man zur sogenannten FERMI-Statistik. Danach ist die Besetzungswahrscheinlichkeit für einen einfachen Quantenzustand der Energie E

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \qquad \text{(Gl. 4.17)}$$

Diese Verteilungsfunktion wird charakterisiert durch die Größe E_F, die sogenannte FERMI-Energie oder FERMI-Kante. Die Funktion $f(E)$ ist in Abb. 4.1 graphisch dargestellt. Für $E \gg E_F$ geht sie gegen Null, für $E \ll E_F$ gegen 1; für $E = E_F$ ist $f(E) = 1/2$. Der Übergang zwischen diesen Werten ist umso abrupter je tiefer die Temperatur ist. Bei $T = 0$ ist $f(E)$ für alle $E > E_F$ gleich 0 und für alle $E < E_F$ gleich 1.

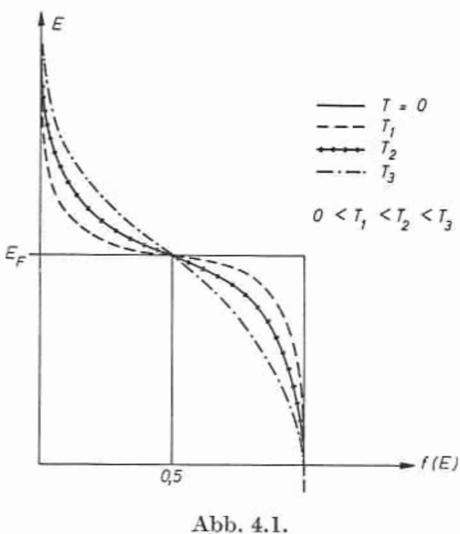


Abb. 4.1.
Die FERMI-Funktion $f(E)$.

Verteilen sich nun n Elektronen auf eine Anzahl Zustände, z. B. ein Energieband, so sind bei $T = 0$ alle Zustände unterhalb $E = E_F$ besetzt, alle darüber leer. Somit ist E_F bestimmt durch die Zahl der Elektronen n (pro cm³) und die Dichte der Zustände auf der Energieachse $N(E)$ (pro cm³ und Energieintervall). Gl.4.18 ist also eine Bestimmungsgleichung für E_F.

$$n = 2 \int_0^\infty N(E)\, f(E, E_F)\, dE \qquad \text{(Gl. 4.18)}$$

Der Faktor 2 kommt hier von der doppelten Besetzungsmöglichkeit jedes Quantenzustandes auf Grund des verschiedenen Spins.

Für $E - E_F \gg kT$, also für schwachbesetzte Zustände, kann Gl. 4.17 durch die BOLTZMANN-Formel

$$f(E) = e^{-(E-E_F)/kT} \qquad \text{(Gl. 4.19)}$$

angenähert werden. Entsprechend kann für fast vollbesetzte Zustände eine FERMI-Funktion für Löcher

$$f_p(E) \equiv 1 - f(E) \qquad \text{(Gl. 4.20)}$$

definiert und im Gebiet $E_F - E \gg kT$ durch

$$f_p(E) = e^{-(E_F-E)/kT} \qquad \text{(Gl. 4.21)}$$

angenähert werden. Diese Näherungen sind im reinen Halbleiter in der Regel gerechtfertigt, im Gegensatz zu sehr hochdotierten Halbleitern, wie sie z. B. in Tunneldioden verwendet werden, und insbesondere im Gegensatz zu Metal-

len. In diesen letzteren Materialien sind die Besetzungswahrscheinlichkeiten zu hoch, um mit Gl. 4.19 oder Gl. 4.21 beschrieben werden zu können. Wir sprechen dann von Entartung bzw. von einem entarteten Elektronengas, während im reinen Halbleiter das Elektronengas als nichtentartet oder „klassisch" bezeichnet wird.

4.1.3 Zustandsdichten und Entartung. Zur genaueren Definition der Grenze der Entartung und zur Berechnung der Massenwirkungskonstanten soll hier ein spezielles Modell eines leitenden Körpers näher betrachtet werden. Bei der ersten quantenmechanischen Formulierung der Elektronentheorie der Metalle hat dieses eine Rolle gespielt [24]. Es überrascht durch seine Einfachheit und Leistungsfähigkeit. Die folgenden Rechnungen können aber ebensogut auf das Bändermodell des Kapitels 3 angewandt werden.

Wir fragen nach der Energieverteilung von Elektronen in einem Volumen V, das die Temperatur T hat. Wir suchen also die Zahl der Elektronen pro cm^3 im Energieintervall von E bis $E + dE$ oder die Dichte der Energiezustände $N(E)$ pro cm^3 und pro Energieintervall. Die Elektronen seien frei, d. h. sie unterliegen keinen elektrischen Feldern, oder sie seien quasi-frei, d. h. ihre Anziehung durch die positiven Ionen des Festkörpergitters und die Abstoßung durch ihresgleichen sei durch den Ansatz einer bestimmten effektiven Masse in Rechnung gestellt worden. Stoßprozesse dieser Elektronen mit dem Gitter, die z. B. zur Berechnung des elektrischen Widerstandes behandelt wurden, bleiben hier außer Betracht. Die das Volumen V begrenzenden Wände sind durch steil ansteigende elektronische Potentiale zu realisieren.

Die Elektronen sind durch die Ladung e, die Masse m_{eff}, ihren Ort (x, y, z) und ihren (Kristall-)Impuls (P_x, P_y, P_z) charakterisiert. Nach der Quantentheorie ist die Festlegung von Ort *und* Impuls nur im Bereich der Unschärferelation gestattet; d. h. ist der Ort bis auf die Unschärfe Δx, Δy, Δz genau bestimmt, so kann der Impuls nur bis auf ΔP_x, ΔP_y, ΔP_z genau bestimmt als werden, und es ist

$$\Delta x \, \Delta y \, \Delta z \, \Delta P_x \, \Delta P_y \, \Delta P_z \geq h^3 \qquad \text{(Gl. 4.22)}$$

Gl. 4.22 berücksichtigt im Partikelbild die Wellennatur der Elektronen. Die Unschärfebedingung läßt sich auch folgendermaßen formulieren: Teilt man den sechsdimensionalen Phasenraum (x, y, z, P_x, P_y, P_z) in Zellen der Größe h^3, so liegt ein Zustand immer innerhalb einer solchen Zelle; mehr zu sagen, ist physikalisch nicht sinnvoll.

Als Ortsunschärfe lassen wir das gesamte Kristallvolumen V zu. Da keine zusätzlichen elektrischen Kräfte auf die Leitfähigkeitselektronen wirken sollen, ist das sinnvoll. Für die Impulsschärfe gilt also (siehe auch Anmerkung am Fuß nächsten der Seite)

$$\Delta P_x \, \Delta P_y \, \Delta P_z = \frac{h^3}{V} \qquad \text{(Gl. 4.23)}$$

Der Impulsteil des Phasenraumes wird zweckmäßigerweise mit Polarkoordinaten vermessen. Eine räumliche Vorzugsrichtung liegt nicht vor, so daß unsere Unschärfebeziehung lautet

$$4\pi\, P^2\, \Delta P = \frac{h^3}{V} \qquad \text{(Gl. 4.24)}$$

Das bedeutet, daß in jeder Kugelschale der Dicke $\Delta P = h^3/(4\pi P^2 V)$ im Mittel ein Quantenzustand liegt. Wir betrachten nun ein Energieintervall dE; es enthält pro cm³ $N(E)\,dE$ Zustände. Dasselbe Intervall kann auch als Impulsintervall dP aufgefaßt werden, das pro cm³ $N(P)\,dP$ Zustände enthält. Hier ist $N(P)$ die Dichte der Zustände pro cm³ und Impulsintervall. Da es sich tatsächlich um dasselbe Intervall handelt, ist die Zahl der darin enthaltenen Zustände in beiden Beschreibungen gleich, also

$$N(E)\,dE = N(P)\,dP.$$

Diese Zahl der Zustände ist aber gleich dem Impulsintervall dP, geteilt durch die Zellengröße ΔP und durch das Gesamtvolumen:

$$N(E)\,dE = \frac{dP}{\Delta P \cdot V} = \frac{4\pi\, P^2\, dP}{h^3} \qquad \text{(Gl. 4.25)}$$

In dieser Gleichung soll nun noch $P^2\,dP$ durch den entsprechenden Energieausdruck ersetzt werden.

Bei der Beschränkung auf freie Elektronen oder auf quasi-freie Elektronen (und Löcher) in der Nähe der Bandränder lautet die klassische Energiegleichung bzw. Gl. 3.47 (mit $k_0 = 0$):

$$E - E_0 = \frac{P^2}{2\,m_{\text{eff}}} \qquad \text{(Gl. 4.27)}$$

wo E_0 die potentielle Energie oder die Energie an der Bandkante, also E_C bzw. E_V ist.

$$dE = \frac{1}{m_{\text{eff}}}\, P\,dP \qquad \text{(Gl. 4.28)}$$

Damit erhält man schließlich die Zustandsdichte $N(E)$ eines freien oder quasifreien Elektrons, d. h. die Zustandsdichte in der Nähe der Bandkante zu

$$N(E)\,dE = \frac{2\,\pi}{h^3}\,(2\,m_{\text{eff}})^{\frac{3}{2}}\,(E - E_0)^{\frac{1}{2}}\,dE \qquad \text{(Gl. 4.29)}$$

Anmerkung:

Diese Bedingung läßt sich aus Gl. 3.31 ableiten: Bei einem Grundgebiet der Länge G, d. h. einem Volumen $V = G^3$, sind die Abstände der Zustände auf der k-Achse $\Delta k = 2\pi/G$. Somit ist $\Delta P_x\, \Delta P_y\, \Delta P_z = \hbar^3\, \Delta k_x\, \Delta k_y\, \Delta k_z = h^3/V$.

Dies kann zur Ermittlung von E_F zusammen mit Gl. 4.17 in Gl. 4.18 eingesetzt werden.

Wir wollen jetzt einen nichtentarteten, reinen n-Halbleiter betrachten, für den Gl. 4.19 gilt. Dann muß aber $E - E_F \gg kT$ sein, wobei die Werte E alle Energien des Leitungsbandes sein können. Es muß also E_F tiefer als das tiefste Niveau des Leitungsbandes sein; die FERMI-Kante muß demnach für den nicht entarteten Halbleiter in der Energielücke zwischen den Bändern liegen.

Dann wird aus Gl. 4.18, 4.19 und 4.29 unter Verwendung des bekannten bestimmten Integrals $\int_0^\infty \sqrt{x}\; e^{-x}\, dx = \frac{\sqrt{\pi}}{2}$

$$n = 2 \int_{E_C}^{\infty} N(E)\, e^{-(E-E_F)/kT}\, dE = N_C\, e^{-(E_C-E_F)/kT} \qquad \text{(Gl. 4.30)}$$

mit der Abkürzung

$$N_C \equiv \frac{2}{h^3} (2\pi m_{\text{eff}}\, kT)^{\frac{3}{2}} \qquad \text{(Gl. 4.31)}$$

die wir die effektive Zustandsdichte des Leitungsbandes nennen.

Dieselbe Betrachtung läßt sich auch für quasi-freie Löcher, d. h. für das Valenzband durchführen. Es ist dann

$$N_p(E)dE = \frac{2\pi}{h^3} (2\, m_{\text{effp}})^{\frac{3}{2}}\, (E_V - E)^{\frac{1}{2}}\, dE$$

und

$$p = 2 \int_{-\infty}^{E_V} N(E)\, f_p(E)\, dE = N_V\, e^{-(E_F-E_V)/kT} \qquad \text{(Gl. 4.32)}$$

mit

$$N_V \equiv \frac{2}{h^3} (2\pi\, m_{\text{effp}} kT)^{\frac{3}{2}} \qquad \text{(Gl. 4.33)}$$

Aus Gl. 4.30–4.33 geht demnach auch hervor, von welchen Dichten n bzw. p ab der Halbleiter entartet ist, nämlich von $n \approx N_C$ und von $p \approx N_V$ ab. Tab. 3 gibt beide Werte für Germanium und Silizium.

Tabelle 3

Effektive Zustandsdichten bei 300° K

	N_C	N_V
Ge	$1{,}04 \cdot 10^{19}\ \text{cm}^{-3}$	$6{,}10 \cdot 10^{18}\ \text{cm}^{-3}$
Si	$2{,}8 \cdot 10^{19}\ \text{cm}^{-3}$	$1{,}02 \cdot 10^{19}\ \text{cm}^{-3}$

Wir sind nun in der Lage, die Massenwirkungskonstante von Gl. 4.1 und 4.6, nämlich das Quadrat der Eigenleitungsdichte, anzuschreiben. Nach Gl. 4.30 und 4.32 ist

$$p\,n = n_i^2 = N_V\,N_C\,e^{-E_G/kT} \qquad \text{(Gl. 4.34)}$$

Weiterhin wollen wir die FERMI-Statistik auf die Besetzung der Störstellen anwenden, um F_2 und F_3 von Gl. 4.9 und 4.10 zu ermitteln. Zu diesem Zweck muß jedoch die Gl. 4.17 etwas abgeändert werden. Wir haben in einem Donator nicht mehr ein einfaches Niveau vorliegen, d. h. für jede Spinrichtung eines, sondern es kann nur ein Elektron angelagert werden, wobei es gleichgültig ist, welchen Spin es hat. Für diesen Fall wird die FERMI-Funktion des Donators [25]:

$$f_D(E) = \frac{1}{1 + \frac{1}{2}\,e^{(E-E_F)/kT}} \qquad \text{(Gl. 4.35)}$$

Und damit wird die Dichte der vollen, nicht-ionisierten Donatoren

$$N_D{}^{\times} = \frac{N_D}{1 + \frac{1}{2}\,e^{(E_C-E_D-E_F)/kT}} \qquad \text{(Gl. 4.36)}$$

Wir setzen dies zusammen mit Gl. 4.30 in Gl. 4.9 unter Berücksichtigung von Gl. 4.7 ein und erhalten ohne weitere Vernachlässigungen:

$$F_2(T) = \frac{N_D{}^{\times}}{(N_D - N_D{}^{\times})\,n} = \frac{2}{N_C}\,e^{E_D/kT} \qquad \text{(Gl. 4.37)}$$

Ganz entsprechend findet man für $F_3(T)$

$$F_3(T) = \frac{N_A{}^{\times}}{(N_A - N_A{}^{\times})\,p} = \frac{2}{N_V}\,e^{E_A/kT} \qquad \text{(Gl. 4.38)}$$

Mit dem Einsetzen dieser Werte in Gl. 4.12 und Gl. 4.6 ist unser Problem der Berechnung der Trägerdichten prinzipiell gelöst. Aus den Trägerdichten n oder p kann dann rückwärts vermöge Gl. 4.30 bzw. 4.32 die Lage der FERMI-Energie berechnet werden.

Man kann nun auch den in Abschnitt 2.4.2.3 angeschnittenen Sonderfall des Reserve-Störleiters (tiefe Temperaturen) anschreiben: Setzt man nämlich Gl. 4.37 in Gl. 4.13 ein, so erhält man

$$n^2 = (N_D - n)\,\frac{N_C}{2}\,e^{-E_D/kT} \qquad \text{(Gl. 4.39)}$$

Dies gibt für tiefe Temperaturen wegen $n \ll N_D$

$$n = \left(\frac{N_D\,N_C}{2}\right)^{\frac{1}{2}} e^{-E_D/2kT} \qquad \text{(Gl. 4.40)}$$

4.1.4 Die isotherme Elektronenatmosphäre. Um die Verhältnisse in Kapitel 5 vorzubereiten, wollen wir nun die Elektronen (oder Löcher) eines Halbleiterkristalls einem konstanten elektrischen Feld aussetzen, das wir von außen anlegen, z. B. indem wir den Kristall isoliert zwischen die Platten eines Kondensators bringen. Dieses Feld F übt auf die Elektronen die Kraft

$$\boldsymbol{K} = -e\boldsymbol{F} \qquad \text{(Gl. 4.41)}$$

aus. Das zugehörige elektrische Potential $\varphi(x, y, z)$ ist definiert durch

$$\boldsymbol{F} = -\nabla\varphi \equiv -\operatorname{grad}\varphi \qquad \text{(Gl. 4.42)}$$

Die potentielle Energie des Elektrons ist dann an einem beliebigen Punkt im Kristall

$$U(x,y,z) = -\int_0^{(x,y,z)} \boldsymbol{K}\, d\boldsymbol{r} = -e\varphi(x,y,z)$$

Die Gesamtenergie E eines Elektrons ist die Summe aus U und aus der kinetischen Energie. Wir beschränken uns auf den Fall einer räumlich konstanten Temperatur (isothermer Fall), wodurch die kinetische Energie konstant wird, und auf einen nicht entarteten Halbleiter, für den die BOLTZMANN-Statistik gilt, und erhalten die räumliche Verteilung der Elektronendichte

$$n(x,y,z) \sim e^{-E/kT} \sim e^{-U(x,y,z)/kT} \sim e^{e\varphi(x,y,z)/kT} \qquad \text{(Gl. 4.43)}$$

Dieses Ergebnis ist wegen Gl. 4.19 bzw. 4.30 gleichbedeutend mit der Aussage, daß die FERMI-Grenze E_F unabhängig vom Ort ist. Das kann allgemein für den Gleichgewichtsfall bewiesen werden, was jedoch hier nicht durchgeführt werden soll. Für Löcher ergibt sich entsprechend

$$p(x,y,z) \sim e^{-e\varphi(x,y,z)/kT} \qquad \text{(Gl. 4.44)}$$

Das Feld F ruft also vermöge Gl. 4.42 und 4.43 eine inhomogene Elektronenverteilung hervor. Diese wiederum verzerrt auf Grund der elektrostatischen Wirkung der Elektronen das Feld, was hier zunächst vernachlässigt werden soll; d. h. das von außen angelegte Feld soll hinreichend groß und die Elektronendichte hinreichend klein sein. Dann ist nach Gl. 4.42

$$F \approx \text{const}; \quad \varphi \approx -Fx \qquad \text{(Gl. 4.45)}$$

wenn F in x-Richtung angenommen wird, und nach Gl. 4.43

$$n \sim e^{-eFx/kT} \qquad \text{(Gl. 4.46)}$$

Diese Gleichung entspricht der barometrischen Höhenformel, was den Ausdruck „isotherme Elektronenatmosphäre" verständlich macht. Ebenso wie bei der Gasatmosphäre der Erde steigt also die Dichte in Richtung des Feldes

(bzw. des Schwerefeldes) exponentiell an. Im Gegensatz zur Erdatmosphäre ändern sich aber die Dichten in der Elektronenatmosphäre, z. B. in *pn*-Sperrschichten, auf sehr kleinen Abständen sehr stark. Die dort vorkommenden Felder von 10^4 V/cm ergeben schon auf $1\,\mu$ bei Zimmertemperatur etwa eine Dichteänderung von zwei Größenordnungen.

4.1.5 Kräftespiel, Nernst-Townsend-Einstein-Beziehung. Es sollen nun die Kräfte erörtert werden, die unter den oben beschriebenen Umständen auf die Elektronen wirken. Infolge des elektrischen Feldes wirkt eine Kraft $-e\boldsymbol{F}$ auf die Elektronen, die eine Elektronenstromdichte in Richtung negativer x erzeugt (für Elektronen ist $q < 0$). In der üblichen Normierung ergibt das eine positive Stromdichte $\boldsymbol{j}_F = \sigma \boldsymbol{F}$ in Richtung positiver x. Bei der Bewegung der Elektronen hat man jetzt zu berücksichtigen, daß die Leitfähigkeitsträger sich nicht im Vakuum befinden, sondern bei der Bewegung im Festkörper Reibungskräften ausgesetzt sind, die zu der angegebenen Formel für die Stromdichte führen. Das Dichtegefälle der Elektronen verursacht außerdem einen Diffusionsstrom $\boldsymbol{j}_D'$ der Teilchen gemäß

$$\boldsymbol{j}_D' = -D_n \nabla n \qquad \text{(Gl. 4.47)}$$

D_n ist die Diffusionskonstante der Elektronen. Der elektrische Strom dieses Diffusionsstromes ist für Elektronen

$$\boldsymbol{j}_D = -e\,\boldsymbol{j}_D' \qquad \text{(Gl. 4.48)}$$

Die resultierende gesamte Stromdichte ist

$$\boldsymbol{j}_n = \boldsymbol{j}_F + \boldsymbol{j}_D = \sigma \boldsymbol{F} + e D_n \nabla n$$

oder unter Berücksichtigung von Gl. 2.3 und Gl. 4.42

$$\boldsymbol{j}_n = -\,e\,n\,\mu_n \nabla\varphi + e D_n \nabla n \qquad \text{(Gl. 4.49)}$$

Für Löcher ergibt sich in entsprechender Schlußweise

$$\boldsymbol{j}_p = -\,e\,p\,\mu_p \nabla\varphi - e D_p \nabla p \qquad \text{(Gl. 4.50)}$$

Im Gleichgewicht muß $\boldsymbol{j} = 0$ sein, d. h. es heben sich Feld- und Diffusionsstrom gerade auf:

$$\mu_n n \nabla\varphi = D_n \nabla n \qquad \text{(Gl. 4.51)}$$

bzw.

$$\mu_p p \nabla\varphi = -\,D_p \nabla p \qquad \text{(Gl. 4.52)}$$

Außerdem gelten bei Gleichgewicht die Gl. 4.43 bzw. 4.44:

$$\nabla n = \frac{e}{kT}\, n \nabla\varphi$$

bzw.

$$\nabla p = -\,\frac{e}{kT}\, p \nabla\varphi$$

und daraus mit Gl. 4.51 bzw. 4.52

$$D_n = \frac{kT}{e} \mu_n = U_T \mu_n \tag{Gl. 4.53}$$

bzw.

$$D_p = \frac{kT}{e} \mu_p = U_T \mu_P \tag{Gl. 4.54}$$

Dabei wurde die Abkürzung „Voltäquivalent der Temperatur" $U_T \equiv kT/e$ benützt (bei Zimmertemperatur = 25 mV). Diese Beziehung zwischen Diffusionskonstante und Beweglichkeit ist für uns von außerordentlicher Wichtigkeit. Sie wurde von NERNST, TOWNSEND und EINSTEIN angegeben. Tabelle 4 enthält diese Größen bei Zimmertemperatur für Germanium und Silizium.

Tabelle 4

Beweglichkeit und Diffusionskonstante

		$\mu \frac{\text{cm}^2}{V\,\text{sec}}$	$D \frac{\text{cm}^2}{\text{sec}}$
Ge	n	3900	97
	p	1900	48
Si	n	1350	34
	p	480	12

4.2 Nicht-Gleichgewicht

4.2.1 Die Bilanzgleichungen. Die bisherigen Betrachtungen des Kapitels 4 bezogen sich auf den Gleichgewichtszustand. Es war bisher die Generation von Elektronen aus dem Valenzband bzw. aus den Störniveaus gleich der Rekombination. Ebenso waren alle Ströme gleich Null. Im letzten Abschnitt wurden zwar die Stromgleichungen 4.49 und 4.50 aufgestellt, aber es wurde lediglich der Fall $\boldsymbol{j} = 0$ näher betrachtet.

Wir wollen uns von jetzt ab auf vollständige Ionisation der Störstellen, d. h. auf $T > 100°\,K$ beschränken, betrachten nun aber den allgemeineren Fall, wo p und n Funktionen des Ortes und der Zeit sind. Dann ist die zeitliche Zunahme der Zahl der Elektronen in einem Volumenelement δV

$$\frac{\partial n}{\partial t} \delta V = (g - r)\, \delta V + \oint \frac{\boldsymbol{j}_n}{e}\, df$$

g ist die Zahl der pro cm³ und sec aus dem Valenzband generierten, r die der rekombinierten Elektronen; $\boldsymbol{j}_n$ ist die von Elektronen getragene elektrische Stromdichte, $-\boldsymbol{j}_n/e$ also die Elektronen-Teilchenstromdichte. Das Integral soll sich über die Oberfläche von δV erstrecken. Damit ist die Zunahme von n

in δV als Netto-Generation, $g - r$, weniger den in der Zeiteinheit aus δV herausfließenden Elektronen dargestellt. Wir wenden den GAUSSschen Satz an und dividieren durch δV:

$$\frac{\partial n}{\partial t} = g - r + \frac{1}{e} \operatorname{div} \boldsymbol{j}_n \qquad \text{(Gl. 4.55)}$$

Entsprechend erhält man für die Löcher

$$\frac{\partial p}{\partial t} = g - r - \frac{1}{e} \operatorname{div} \boldsymbol{j}_p \qquad \text{(Gl. 4.56)}$$

Diese Gleichungen nennen wir die Bilanzgleichungen. Sie bestimmen zusammen mit den Stromgleichungen 4.49 und 4.50 und mit der weiter unten angeführten POISSON-Gleichung (Gl. 4.63) die Vorgänge in fast allen Halbleiterbauelementen, insbesondere in Gleichrichtern und Transistoren. Wir wollen die Bilanzgleichungen in den folgenden Abschnitten anwenden.

4.2.2 Trägerdichte bei schnellen Feldänderungen. Wir haben in Abschnitt 4.1.4 die Gleichgewichtsdichteverteilung für Elektronen in einem zeitlich konstanten Feld betrachtet. Wir haben festgestellt, daß diese Gleichgewichtsverteilung zustande kommt durch ein dynamisches Gleichgewicht zwischen dem Feldstrom j_F und dem Diffusionsstrom j_D. Wird das elektrische Feld plötzlich geändert, so wird sich eine neue Dichteverteilung einstellen. Wir wollen nun abschätzen, wie lange die Elektronenatmosphäre braucht, um diese neue Verteilung anzunehmen. Hierzu verwenden wir die Bilanzgleichung 4.55, wobei wir wegen ihrer relativen Kleinheit die Generation und die Rekombination vernachlässigen können. Außerdem betrachten wir das Problem in seiner eindimensionalen Form und ersetzen den Divergenzausdruck durch die Ableitungen nach der Ortskoordinate x und schreiben

$$\frac{\partial n}{\partial t} = \frac{1}{e}\left(\frac{\partial j_F}{\partial x} + \frac{\partial j_D}{\partial x}\right) \qquad \text{(Gl. 4.56)}$$

Da es uns nur darum geht, die Größenordnung der Einstellzeit zu ermitteln, genügt es, einen besonders einfachen Fall zu behandeln. Das Feld zur Zeit $t \leq 0$ habe die Elektronen so stark nach einer Seite des Kristalls gedrängt, daß wir n als Zackenfunktion an der Stelle $x = 0$ ansehen können, d. h. alle Elektronen sollen zu Anfang bei $x = 0$ sein. Nun soll zur Zeit $t = 0$ das Feld ganz abgeschaltet werden, und es soll verfolgt werden, wie sich $n(x, t)$ auf das neue Feld, und zwar $F = 0$, einstellt.

Da für $F = 0$ auch $j_F = 0$ ist, vereinfacht sich Gl. 4.56 unter Berücksichtigung von Gl. 4.47 und 4.48 zu:

$$\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} \qquad \text{(Gl. 4.57)}$$

Dies ist die FICKsche Diffusionsgleichung. Wir rechnen also so, als ob der

Ladungsausgleich allein durch die Diffusion erfolgen würde. Tatsächlich stoßen sich die bei $x = 0$ versammelten Elektronen elektrostatisch ab, d. h. die Annahme $F = 0$ stimmt eben nicht, auch wenn das äußere Feld abgeschaltet ist. Dies führt dazu, daß die Ladungsanhäufung bei $x = 0$ noch schneller ausgeglichen wird als durch Diffusion allein. Die Berechnung nach Gl. 4.57 wird also eine obere Grenze der Einstellzeit liefern.

Für die obige Anfangsbedingung ist die Lösung unserer Differentialgleichung bekannt und ist gegeben durch die GAUSSsche Glockenkurve

$$n(x, t) = \text{const}\; e^{-x^2/4D_n t} \qquad \text{(Gl. 4.58)}$$

Unsere Frage läßt sich nun folgendermaßen präzisieren: Nach welcher Zeit ist der Konzentrationsunterschied zwischen $x = 0$ und $x = x_1$ gerade gleich dem Faktor e? Die Antwort ist: Wenn der Exponent gleich 1 ist, also $x_1^2/(4\,D_n\,t) = 1$ oder $t = x_1^2/(4\,D_n)$. Wir interessieren uns speziell für den Trägerdichtenausgleich in Feldzonen, wie sie in *pn*-Übergängen in Gleichrichtern und Transistoren vorkommen. Solche Schichten sind von der Größenordnung $1\,\mu$. Wir erhalten somit mit $x_1 = 10^{-4}$ cm für Elektronen in Germanium (siehe Tab. 4) eine Einstellzeit $t = 2{,}5 \cdot 10^{-11}$ sec. Mit anderen Worten, die Elektronenatmosphäre in einem *pn*-Übergang kann Wechselspannungen bis zu mehreren GHz folgen.

4.2.3 Einstellung der Ladungsneutralität im Halbleiter. Im obigen Problem haben wir die Wirkung des Feldes, das von den bei $x = 0$ angehäuften Elektronen verursacht wird, vernachlässigt. Wir wollen nun einen Fall behandeln, wo wir umgekehrt das Diffusionsglied vernachlässigen.

Wird an irgendeiner Stelle eines Halbleiterkristalls eine kleine elektrische Ladungsmenge zugefügt, z. B. durch kurzzeitige Injektion aus einem Kontakt oder durch den Beschuß mit einem Elektronen- oder Ionenstrahl, so wird sich diese Ladung in sehr kurzer Zeit auf den ganzen Kristall verteilen und, falls eine leitende Verbindung zur Erde besteht, durch diese abfließen. Es soll nun berechnet werden, wie schnell sich diese Ladungsgleichheit bzw. im Falle der leitenden Verbindung die Ladungsneutralität wieder einstellt.

Wir gehen von der Kontinuitätsgleichung aus, die allgemein die Erhaltung der elektrischen Ladung ausdrückt.

$$\frac{\partial \varrho}{\partial t} = -\operatorname{div} \boldsymbol{j} \qquad \text{(Gl. 4.59)}$$

ϱ ist die Ladungsdichte. Wird diese als Funktion der Zahl der zugefügten

Löcher pro Volumeneinheit δp und der Elektronen δn ausgedrückt, so ist

$$\varrho = e\,(\delta p - \delta n) \qquad \text{(Gl. 4.60)}$$

und

$$\frac{\partial \varrho}{\partial t} = e\,\frac{\partial \delta p}{\partial t} - e\,\frac{\partial \delta n}{\partial t}$$

Gl. 4.59 kann auch aus unseren Bilanzgleichungen 4.55 und 4.56 gewonnen werden durch Multiplikation mit e und Subtraktion beider Gleichungen. Es ist außerdem nach Gl. 4.49 und 4.50

$$\boldsymbol{j} = \boldsymbol{j}_p + \boldsymbol{j}_n = (e\mu_p\,p + e\mu_n\,n)\,\boldsymbol{F} + e\,D_n \nabla\delta n - e\,D_p \nabla\delta p \qquad \text{(Gl. 4.61)}$$

Für kleine zugefügte Ladungsdichten sind die beiden letzten Glieder klein gegenüber dem ersten. Die Klammer ist gleich der Leitfähigkeit $\sigma_p + \sigma_n = \sigma$, und diese ist nicht sehr verschieden vom Gleichgewichtswert σ_0. Man erhält durch Einsetzen in Gl. 4.59

$$\frac{\partial \varrho}{\partial t} = -\,\sigma_0\,\mathrm{div}\,\boldsymbol{F} \qquad \text{(Gl. 4.62)}$$

Weiterhin benötigt man die Poisson-Gleichung. Sie verknüpft div $\boldsymbol{F}$ mit der Ladungsdichte ϱ:

$$\mathrm{div}\,\boldsymbol{F} = -\,\Delta\,\varphi = \frac{\varrho}{\varepsilon\,\varepsilon_0} \qquad \text{(Gl. 4.63)}$$

wo ε die relative Dielektrizitätskonstante und ε_0 die absolute des Vakuums ist. Mit Gl. 4.62 erhält man eine Differentialgleichung für ϱ

$$\frac{\partial \varrho}{\partial t} = -\,\frac{\sigma_0}{\varepsilon\,\varepsilon_0}\,\varrho \qquad \text{(Gl. 4.64)}$$

mit der Lösung

$$\varrho \sim e^{-\frac{\sigma_0}{\varepsilon\,\varepsilon_0}\,t} \qquad \text{(Gl. 4.65)}$$

Dies bedeutet, daß die injizierte Zusatzladung in einer Zeit von der Größenordnung $\varepsilon\,\varepsilon_0/\sigma_0$ zerfließt. Das ist für 10 Ωcm ($\sigma_0 = 0{,}1$) und Ge ($\varepsilon = 16$) eine Zeit von $1{,}4 \cdot 10^{-11}$ sec. Man kann also sagen, daß sich im Halbleiter die Ladungsneutralität fast unmittelbar einstellt.

Dieses Zerfließen der Ladung kann nun so erfolgen, daß die zugefügte Elektronenwolke δn selbst zerfließt vermöge des Gliedes $-\mathrm{div}\,\boldsymbol{j}_n$, also als Elektronenteilchenstrom aus dem Ladungsgebiet heraus. Es kann aber auch erfolgen durch Hereinfließen von Löchern δp, d. h. durch den Anteil $-\mathrm{div}\,\boldsymbol{j}_p$. Im letzteren Falle entsteht aus der Elektronenwolke eine Wolke aus einer gleichen Zahl von Elektronen und Löchern, die selbst elektrisch neutral ist ($\delta n = \delta p$). Die relative Größe dieser beiden Ausgleichsprozesse kann aus den Bilanzgleichungen berechnet werden, wenn $g - r$ vernachlässigt wird. Es

folgt aus Gl. 4.55, 4.56 und den Vernachlässigungen im Zusammenhang mit Gl. 4.62

$$\frac{\partial p}{\partial t} \approx -\frac{1}{e}\,\sigma_p \operatorname{div} \boldsymbol{F} \qquad\qquad -\frac{\partial n}{\partial t} \approx -\frac{1}{e}\,\sigma_n \operatorname{div} \boldsymbol{F}$$

und daraus das Verhältnis der Löcherzunahme zur Elektronenabnahme

$$\frac{\partial p/\partial t}{-\partial n/\partial t} = \frac{\sigma_p}{\sigma_n} = \frac{\mu_p\, p}{\mu_n\, n}$$

Aus dieser Gleichung geht hervor, daß eine Elektronenwolke in p-Material ($\sigma_p \gg \sigma_n$) durch Löcherzufluß neutralisiert wird, während sie in n-Material ($\sigma_n \gg \sigma_p$) selbst zerfließt. Entsprechend ist es mit einer Löcherwolke in n- bzw. p-Material. Mit anderen Worten, eine Majoritätenwolke zerfließt, während eine Minoritätenwolke durch Zufluß von Majoritäten neutralisiert wird. Erst im Verlauf einer längeren Zeit verschwindet diese neutrale Wolke dann durch Rekombination. Dieser Rekombinationsprozeß soll im folgenden Abschnitt näher untersucht werden.

4.2.4 Die Lebensdauer von Minoritäten. Wird in einem n-Halbleiter, der im thermischen Gleichgewicht n_N Elektronen und p_N Löcher in der Volumeneinheit enthält, durch irgendeine Maßnahme zur Zeit $t = 0$ in dem betrachteten Volumen gleichmäßig die Löcherdichte um δp erhöht, so erhöht sich nach dem oben Gesagten in sehr kurzer Zeit ebenfalls die Elektronendichte um $\delta n = \delta p$. Es ist also gleichgültig, ob man kurzzeitig Minoritäten zufügt, oder beispielsweise durch kurzzeitige Einstrahlung von Licht geeigneter Wellenlänge Paare von Defektelektronen und Elektronen schafft. Mit diesem Zustand $\delta p = \delta n > 0$ gehen wir jetzt in die Bilanzgleichung ein, berücksichtigen aber jetzt im Gegensatz zu den Abschnitten 4.2.2 und 4.2.3 das Generations- und Rekombinationsglied $g - r$.

Wir nehmen an, daß die Erhöhung der Löcherdichte $\delta p = p - p_N$ überall im Kristall gleich groß ist. Dann ist $\nabla p = 0$. Ist ein eventuell vorhandenes Feld $\boldsymbol{F}$ räumlich konstant, so ist nach Gl. 4.50

$$\operatorname{div} \boldsymbol{j}_p = e\mu_p\, p \operatorname{div} \boldsymbol{F} - e D_p \Delta p = 0 \qquad \text{(Gl. 4.66)}$$

und nach Gl. 4.56

$$\frac{\partial p}{\partial t} = g - r \qquad \text{(Gl. 4.67)}$$

Diese Gleichung beschreibt dann die zeitliche Abnahme der Löcherdichte nach einer kurzzeitigen Erhöhung δp zur Zeit $t = 0$.

Am einfachsten ist der Rekombinationsvorgang, wenn das Elektron direkt vom Leitungsband in ein Loch im Valenzband herunterfällt. Dann ist r proportional der Dichte der Elektronen n und der Dichte der Löcher p, also

$$r = r'\, p\, n \qquad \text{(Gl. 4.68)}$$

Die Generation g ist unabhängig von p und n. Im thermischen Gleichgewicht ist $\partial p/\partial t = 0$, $p = p_N$ und $n = n_N$. Wegen Gl. 4.67 muß deshalb sein:

$$g = r' p_N n_N \qquad \text{(Gl. 4.69)}$$

Wird dies wieder in Gl. 4.67 eingesetzt, so erhält man wegen $\delta p = \delta n$

$$\frac{\partial p}{\partial t} = r' p_N n_N - r'(p_N + \delta p)(n_N + \delta p) = -r'(p_N + n_N)\,\delta p - r'\,\delta p^2$$

Für kleine Minoritätenerhöhung ($\delta p \ll n_N$) und starken n-Typ ($p_N \ll n_N$) vereinfacht sich dies zu

$$\frac{\partial p}{\partial t} = -r' n_N\,\delta p \qquad \text{(Gl. 4.70)}$$

Die Lösung dieser Differentialgleichung ist

$$\delta p(t) = \delta p(0)\; e^{-t/\tau_p} \qquad \text{(Gl. 4.71)}$$

mit

$$\tau_p \equiv \frac{1}{r' n_N} \qquad \text{(Gl. 4.72)}$$

Wir haben also einen exponentiellen Abfall der zugefügten Löcherdichte mit der charakteristischen Zeit τ_p, die wir Löcherlebensdauer nennen. Ganz entsprechend verhält sich eine Erhöhung der Elektronendichte $\delta n = n - n_P$ in p-Material. Es ist

$$\delta n(t) = \delta n(0)\; e^{-t/\tau_n} \qquad \text{(Gl. 4.73)}$$

mit der Elektronenlebensdauer

$$\tau_n \equiv \frac{1}{r' p_P} \qquad \text{(Gl. 4.74)}$$

wo p_P die Gleichgewichtsdichte der Löcher im p-Gebiet ist.

Die Minoritätenlebensdauer τ_p bzw. τ_n hängt nicht nur von der Majoritätendichte, sondern implizit durch r' von allen möglichen Arten von Gitterstörungen ab, wie z. B. von Fehlstellen, Versetzungen (vgl. Abschnitt 2.2.3), nicht dotierenden Fremdstoffen und selbstverständlich von der Temperatur.

Ist die Reaktion zwischen Löchern und Elektronen bei der Rekombination und Generation nicht eine direkte, sondern sind Haftstellen (sogenannte Traps) beteiligt, so hängt r in komplizierterer Weise als in Gl. 4.68 von p und n ab. Man entwickelt dann aber nach p und n an der Stelle p_N, n_N

$$r = r(p_N, n_N) + \frac{\partial r}{\partial p}\,\delta p + \frac{\partial r}{\partial n}\,\delta n + \dots$$

und erhält mit

$$r' \equiv \left(\frac{\partial r}{\partial n} + \frac{\partial r}{\partial p}\right)\frac{1}{n_N}$$

formal dasselbe Ergebnis wie oben, jedoch mit dem Unterschied, daß r' jetzt eine Funktion von p_N und n_N ist (Näheres siehe SHOCKLEY und READ [45] und HOFFMANN [46]).

Ist δp zwar klein, aber nicht unabhängig vom Ort, d. h. ist die Erzeugung der Zusatzminoritäten nicht überall im Kristall gleichgroß, so läßt sich die Bilanzgleichung wieder verallgemeinern durch Einsetzen der Gleichungen 4.68, 4.69 und 4.72 in Gl. 4.56:

$$\frac{\partial p}{\partial t} = -\frac{\delta p}{\tau_p} - \frac{1}{e}\operatorname{div} \boldsymbol{j}_p$$

Außerdem ist es für das Folgende zweckmäßig, noch ein Glied für die zusätzliche Trägergeneration durch Beleuchtung g_L hinzuzufügen. g_L kann dabei eine Funktion des Ortes (bevorzugte Absorption an der beleuchteten Oberfläche) und der Zeit (Wechsellicht, Lichtimpulse) sein und hängt außerdem von der Intensität und Wellenlänge des Lichtes ab, da nur solche Photonen, die eine Energie $> E_G$ haben, Paare erzeugen können. Wir haben damit ($\delta p = p - p_N$):

$$\frac{\partial p}{\partial t} = g_L - \frac{\delta p}{\tau_p} - \frac{1}{e}\operatorname{div} \boldsymbol{j}_p \qquad \text{(Gl. 4.75)}$$

und entsprechend für Elektronen im p-Halbleiter ($\delta n = n - n_P$):

$$\frac{\partial n}{\partial t} = g_L - \frac{\delta n}{\tau_n} + \frac{1}{e}\operatorname{div} \boldsymbol{j}_n \qquad \text{(Gl. 4.76)}$$

4.2.5 Photoleitfähigkeit und Lebensdauermessung. Ein dünner Stab sei aus einem n-Typ-Halbleiterkristall herausgeschnitten, an seinen Enden kontaktiert und über einen Widerstand mit einer Gleichstromquelle verbunden (Abb. 4.2). Von oben werde er gleichmäßig beleuchtet. Das Licht wird in einer sehr dünnen Oberflächenschicht im Halbleiter absorbiert und erzeugt dort Elektronen und Löcher, die nach dem Innern, d.h. in y-Richtung, diffundieren. Die Batterie erzeugt in x-Richtung ein konstantes Feld F. Es ist dann nach Gl. 4.50

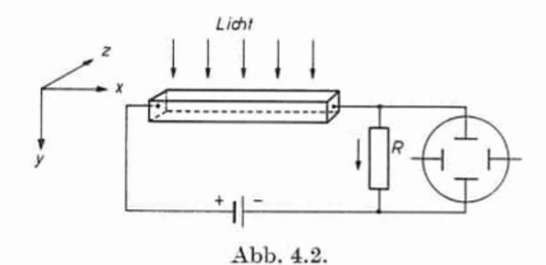


Abb. 4.2.
Messung der Minoritätenlebensdauer aus dem Abklingen der Photoleitfähigkeit.

$$\frac{1}{e}\operatorname{div} \boldsymbol{j}_p = -D_p \frac{\partial^2 p}{\partial y^2} \neq 0 \qquad \text{(Gl. 4.77)}$$

Dies ist streng genommen in Gl. 4.75 einzusetzen, und die entstehende Differentialgleichung für die Oberflächenschicht und das Innere getrennt zu berechnen und aneinander anzupassen. Wir können uns aber auch ohne Rechnung vorstellen, daß sich durch diesen Diffusionsstrom in y-Richtung bei dünnem

Stäbchen sehr rasch ein stationärer Mittelwert von δp einstellen wird, den wir $\overline{\delta p}$ nennen wollen und für den Gl. 4.75 gilt ohne das Stromdivergenzglied

$$\frac{\partial \overline{\delta p}}{\partial t} = \overline{g_L} - \frac{\overline{\delta p}}{\tau_p} \qquad \text{(Gl. 4.78)}$$

Bei Gleichlicht wird $\partial \overline{\delta p}/\partial t = 0$, und es ergibt sich eine Erhöhung der Löcher- und Elektronendichte

$$\overline{\delta p} = \overline{\delta n} = \tau_p \, \overline{g_L}$$

und damit eine Erhöhung der Leitfähigkeit, die sogenannte Photoleitfähigkeit

$$\overline{\delta \sigma} = e \mu_p \overline{\delta p} + e \mu_n \overline{\delta n} = e(\mu_p + \mu_n)\, \tau_p \overline{g_L} \qquad \text{(Gl. 4.79)}$$

Sie kann z. B. als Zunahme des Spannungsabfalls am Widerstand R gemessen werden.

Wird statt Gleichlicht Wechsellicht eingestrahlt, so kann am Oszillographen der Abb. 4.2 eine Wechselspannung beobachtet werden. Wir wollen ganz kurz den Fall einer pulsförmigen Lichteinstrahlung an Hand von Gl. 4.78 betrachten, weil dies eines der wichtigsten Verfahren zur Messung der Trägerlebensdauer ist [26].

Zur Zeit $t = 0$ habe der Lichtimpuls eine mittlere Zusatzträgerdichte $\overline{\delta p(0)}$ erzeugt; das Licht wird jetzt abgeblendet und für die Zeit $t > 0$ ist $\overline{g_L} = 0$. Man hat dann nach Gl. 4.78

$$\frac{\partial \overline{\delta p}}{\partial t} = -\frac{\overline{\delta p}}{\tau_p} \qquad \text{(Gl. 4.80)}$$

mit der Lösung

$$\overline{\delta p(t)} = \overline{\delta p(0)}\, e^{-t/\tau_p} \qquad \text{(Gl. 4.81)}$$

ganz analog zu Gl. 4.71. Es wird also auf dem Oszillographen eine Exponentialkurve aufgezeichnet, aus der man leicht die Lebensdauer τ_p ermitteln kann.

4.2.6 Oberflächenrekombination. Wir haben bisher stillschweigend so getan, als ob die Rekombination r und damit die Lebensdauer τ_p bzw. τ_n vom Ort unabhängig wären. Tatsächlich ist in der Regel die Rekombination an der Oberfläche sehr viel größer als im Innern des Kristalls. Die Oberfläche wirkt dann als eine Senke für den Minoritätenstrom. Die Größe des an der Oberfläche verschwindenden, d. h. rekombinierenden Minoritätenstroms, der senkrecht in die Oberfläche hineinfließt ($j_{p\perp}$; geschrieben für Löcher im n-Material), ist

$$\frac{1}{e}\, j_{p\perp} \equiv s\, \delta p \qquad \text{(Gl. 4.82)}$$

$j_{p\perp}$ ist also proportional zur Überschußdichte $\delta p = p - p_N$ der Minoritäten. s ist die sogenannte Oberflächenrekombinations-Geschwindigkeit (Dimension cm/sec). Sie ist selbstverständlich von der Art und Vorbehandlung der Ober-

fläche abhängig. Sie ist bei gut geätzten Oberflächen sehr klein ($s = 100$ cm/sec), bei geschmirgelten sehr groß ($s = 10000$ cm/sec). Im allgemeinen ist sie bei Silizium größer als bei Germanium.

Diese Oberflächenrekombination beeinflußt nun den Abfall der Minoritätendichte in kleinen und dünnen Halbleiterstrukturen wie z. B. Transistoren sehr entscheidend. Dies sei am Beispiel einer dünnen Platte (Abb. 4.3) mit überall gleicher Oberflächenrekombination klargemacht. Die Platte sei n-Typ, habe eine Dicke $2a$ und sei so groß, daß wir mit einem eindimensionalen Problem (in y-Richtung) rechnen können. Es gilt also Gl. 4.77, die wir in Gl. 4.75 (mit $g_L = 0$) einsetzen:

$$\frac{\partial \delta p}{\partial t} = -\frac{\delta p}{\tau_p} + D_p \frac{\partial^2 \delta p}{\partial y^2} \qquad \text{(Gl. 4.82a)}$$

Abb. 4.3. Zur Berechnung der effektiven Lebensdauer.

Diese Differentialgleichung mit der Randbedingung Gl. 4.82 hat Lösungen der Form

$$\delta p \sim \cos by \cdot e^{-t/\tau_{\mathrm{eff}}} \qquad \text{(Gl. 4.83)}$$

τ_{eff} ist eine effektive Lebensdauer, die, wie wir sehen werden, von τ_p und s abhängt. In Gl. 4.83 können die beiden Konstanten b und τ_{eff} verschiedene Werte haben, und statt Cosinus kann Sinus stehen. Die Lösung mit dem größten Wert von τ_{eff} ist diejenige, die uns interessiert, weil alle Lösungen höherer Ordnung schneller gegen Null gehen und nur diese schließlich übrig bleibt.

Einsetzen von Gl. 4.83 in 4.82a ergibt

$$\frac{1}{\tau_{\mathrm{eff}}} = \frac{1}{\tau_p} + D_p b^2 \qquad \text{(Gl. 4.84)}$$

und von Gl. 4.83 in 4.82 ergibt mit 4.50

$$D_p b \operatorname{tg}(ba) = s \qquad \text{(Gl. 4.85)}$$

Dies kann für die beiden Grenzfälle $s \to 0$ und $s \to \infty$ explizit aufgelöst und in Gl. 4.84 eingesetzt werden und ergibt:

$$\frac{1}{\tau_{\mathrm{eff}}} = \frac{1}{\tau_p} + \begin{cases} s/a & \text{für } s \to 0 \\ \dfrac{D_p \pi^2}{4a^2} & \text{für } s \to \infty \end{cases} \qquad \text{(Gl. 4.86)}$$

Man sieht also, daß, nachdem die höheren Lösungen abgeklungen sind, die Minoritätendichte gemäß Gl. 4.83 mit einer effektiven Lebensdauer τ_{eff} abnimmt, die kürzer als die Volumenlebensdauer τ_p ist, und die nach Gl. 4.86 von der Oberflächenrekombination und von der Geometrie abhängt.

Fünftes Kapitel

Der PN-Übergang

5.1 Einleitung

Wir haben in den vorangegangenen Kapiteln gesehen, welch entscheidenden Einfluß die Störstellenkonzentrationen, insbesondere diejenigen der Donatoren und Akzeptoren, auf das elektrische Verhalten des Halbleiters haben. Dabei wurde bisher immer stillschweigend angenommen, daß die Störstellenkonzentrationen überall im betrachteten Kristallbereich gleich sind. Ist dies nicht der Fall, so ergeben sich mit der örtlich veränderlichen Störstellendichte Konzentrationsgradienten von Elektronen und Löchern. Diese erzeugen nach Gl. 4.47 Diffusionsströme, welche die Trägerdichteunterschiede einzuebnen trachten. Dadurch ergeben sich jedoch wegen der ortsfesten Donatoren und Akzeptoren elektrische Raumladungen und Felder, die im Gleichgewichtsfall gerade so groß werden, daß der durch die Diffusion hervorgerufene Trägerstrom zum Stillstand kommt (vgl. Abschnitt 4.1.5). Diese allgemeine Betrachtung wird besonders deutlich und ist besonders wichtig beim sogenannten *PN*-Übergang, den wir im folgenden Kapitel ausschließlich behandeln wollen.

Ein solcher *PN*-Übergang entsteht, wenn sich die Donatoren- und Akzeptorendichten im Kristall örtlich so ändern, daß zwei aneinandergrenzende Gebiete verschiedenen Leitungstyps vorliegen. Wir sprechen dann von einem *PN*-Kristall oder auch *PN*-Schichtkristall. In Abb. 5.1 sind verschiedene Arten von *PN*-Übergängen schematisch dargestellt. Der einfachste, reine Fall liegt vor, wenn die *P*-Zone nur Akzeptoren und die *N*-Zone nur Donatoren enthält (a und b). Dieser Übergang ist symmetrisch (a) oder asymmetrisch (b), je nachdem, ob die Donatoren- und Akzeptorendichten auf beiden Seiten gleich oder verschieden sind. In der Regel sind jedoch beide Zonen mit der entgegengesetzten Störstellenart teilweise kompensiert (c), wobei aber der Leitungstyp nach Abschnitt 2.4.2.4 durch die überwiegende Störstellenart bestimmt wird. In den Beispielen a, b und c sind die Störstellendichten innerhalb des *P*- und des *N*-Gebietes konstant und ändern sich am *PN*-Übergang sprungartig, abrupt. Abb. 5.1d gibt ein Beispiel eines flachen graduellen Überganges, wo die Akzeptorendichte N_A von links nach rechts abnimmt, während die Donatorendichte N_D überall konstant ist. Der *PN*-Übergang liegt dann dort, wo $N_A(x) = N_D$ ist.

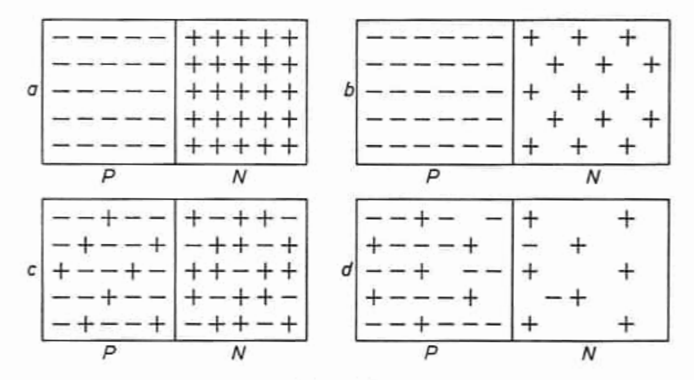


Abb. 5.1.
PN-Übergänge (— bedeuten Akzeptoren, + bedeuten Donatoren).
a) reiner, abrupter symmetrischer PN-Übergang. b) reiner, abrupter unsymmetrischer PN-Übergang. c) abrupter symmetrischer PN-Übergang in teilweise kompensiertem Material. d) flacher PN-Übergang mit konstanter Donatorendichte und von links nach rechts abnehmender Akzeptorendichte.

Das elektrische Verhalten des *PN*-Schichtkristalls ist das Tor zum Verständnis der Physik der Dioden, Gleichrichter, Photozellen und Transistoren, weil diese Bauelemente als wesentliche Bestandteile einen bzw. zwei *PN*-Übergänge enthalten. Darüber hinaus wird der *PN*-Übergang zur Erforschung und Messung vieler Phänomene im Halbleiter verwendet, wie z. B. Messung der Minoritätendichten und der Lebensdauer (siehe Abschnitt 5.5.3), Messung der Beweglichkeit im „Drift"-Versuch [27], Studium der Halbleiteroberfläche [28], Messung der Diffusion von Verunreinigungen im Halbleiter [29, 30] und der Fortpflanzung von Schallquanten (Phonon Drag Effect) [31]. Außerdem ist die Wirkungsweise des *PN*-Übergangs von Wichtigkeit zum Verständnis der Halbleiterthermoelemente und des Halbleiter-PELTIER-Effekts [32].

Beim Germanium-*PN*-Schichtkristall ist es zum ersten Male gelungen, das Gleichrichterverhalten aus dem nach unabhängigen Verfahren gemessenen Halbleitereigenschaften der beiden Halbleiterhälften quantitativ zu berechnen [33, 34, 35]. So wie das Germanium für die Halbleiterphysik als ein Schulstoff aufzufassen ist, ist der Ge-Schichtkristall ein Paradestück der Gleichrichterphysik geworden. Hier sind die einfachen Verhältnisse, die wir in diesem Kapitel zu entwickeln haben, zum größten Teil realisiert, wenn auch die weitere Entwicklung gezeigt hat, daß die Ge-Oberfläche das hier zu entwerfende Bild verändern kann. Auch der dem Germanium sehr ähnliche Si-Gleichrichter zeigt — allerdings bei noch besseren Gleichrichtereigenschaften als Ge — ein komplizierteres Verhalten, das nur durch zusätzliche Annahmen zu deuten ist [9, 36].

Trotz mehr als 30 Jahren Erfahrung ist eine solche Berechnung bei den Trockengleichrichtern aus Kupferoxydul und Selen nie möglich gewesen. Erst die Ergebnisse an Germanium und Silizium haben auch hier ein besseres Ver-

ständnis ermöglicht. Die technologische Entwicklung der Selen- und Kupferoxydulgleichrichter ist übrigens ein gutes Beispiel dafür, welch erstaunliche Erfolge auch ohne viel Grundlagenkenntnis zunächst erzielt werden können. Wenn es dann jedoch darum geht, das Höchste an Qualität herauszuholen, dann ist dasjenige Bauelement im Vorteil, dessen physikalische Grundlagen genau bekannt sind. Dies ist an Hand von Abb. 5.2 am Beispiel des Silizium- und Germanium-Gleichrichters, verglichen mit dem aus Kupferoxydul und Selen, eindringlich sichtbar.

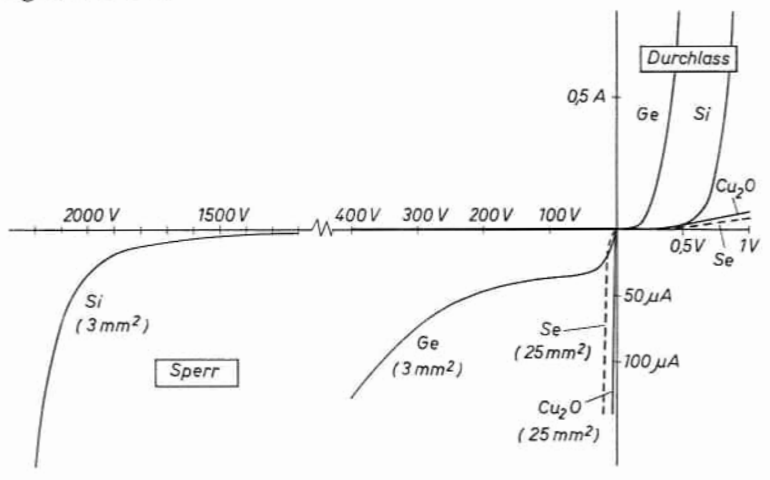


Abb. 5.2.
Vergleich der Kennlinien verschiedener Gleichrichterarten (Fluß- und Sperrkennlinien sind in verschiedenen Maßstäben gezeichnet).

Interessant ist es auch, sich in diesem Zusammenhang daran zu erinnern daß der erste Transistor, der Spitzenkontakttransistor, von Bardeen und Brattain [37] im Laboratorium „entdeckt" wurde, während der überlegene, heute ausschließlich verwendete Flächentransistor ein Jahr später von Shockley [33] mit Hilfe der PN-Theorie am Schreibtisch „erfunden" wurde.

5.2 Charakteristische Beispiele von PN-Kristallen

5.2.1 Silizium-Gleichrichter. Abb. 5.3 zeigt den schematischen Aufbau eines Gleichrichters. Er besitzt einen PN-Übergang und niederohmige, d. h. sperrfreie Kontakte an den Enden. Diese Sperrfreiheit der Kontakte wird durch eine sehr hohe Dotierung der Enden bis zur Entartung erreicht (mit P^+ bzw. N^+ bezeichnet). Dadurch werden eventuell auftretende Sperrschichten so schmal, daß sie von den Elektronen durchtunnelt werden können.

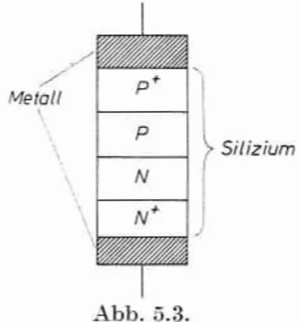
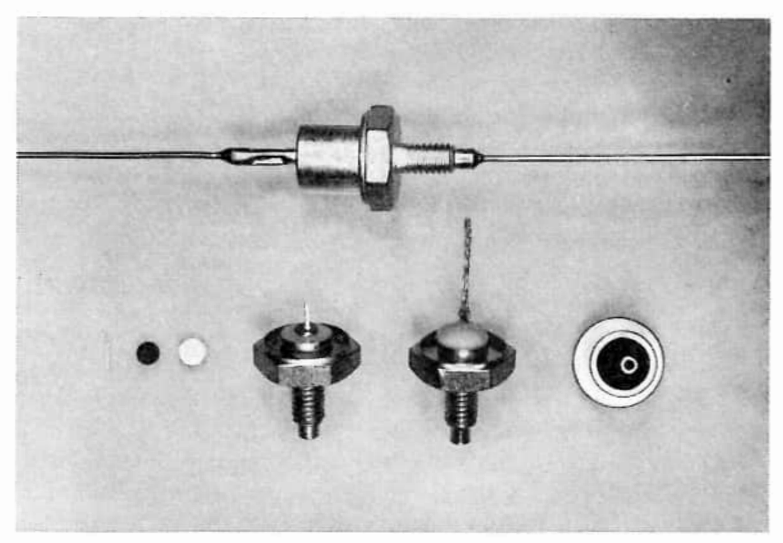


Abb. 5.3.
Schematischer Aufbau eines Siliziumgleichrichters.



Tafel 7.
Einzelteile und Aufbau eines legierten Silizium-1-Ampère-Gleichrichters (Werkphoto Intermetall).

Tafel 7 zeigt die Einzelteile und den Aufbau eines 1-Ampère-Gleichrichters, wie er z. B. zur Spannungsversorgung in Fernsehgeräten verwendet wird. Seine Kennlinie ist die des Siliziumgleichrichters in Abb. 5.2. Durch Anwendung von großflächigeren *PN*-Übergängen und größeren Gehäusen gelangt man zu Leistungsgleichrichtern für höhere Ströme. Die Kennlinie einer 20-Ampère-Type ist in Abb. 5.4 abgebildet.

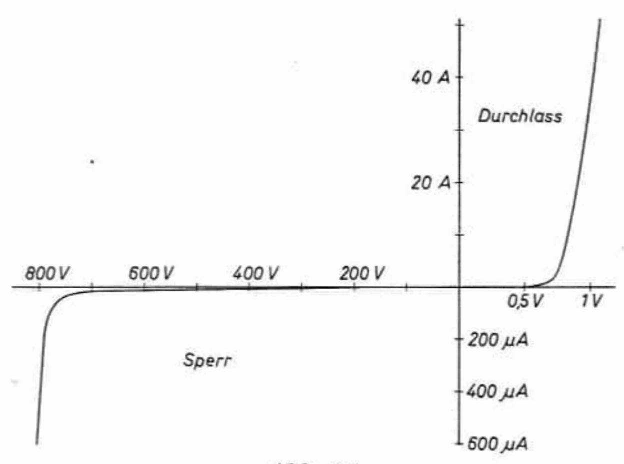


Abb. 5.4.
Kennlinie eines Si-Gleichrichters für 20 Ampère.

5.2.2 Germanium-Punktkontaktdiode. Abb. 5.5 gibt das Schema einer Punktkontaktdiode wieder. Ein Germaniumplättchen vom *n*-Typ ist mit einer Seite auf eine metallische Basis aufgelötet. Auf der Oberseite des Plättchens ist eine Metallspitze aufgesetzt und durch elektrisches Formieren leicht angeschweißt. Dadurch entsteht ein kleinflächiger *PN*-Übergang und die in Abb. 5.6 wiedergegebene Kennlinie. Solche Dioden sind wegen ihrer kleinen Kapazität geeignet zur Hochfrequenzgleichrichtung und -mischung, als Diskriminatoren in UKW-Empfängern, als Modulatoren und Demodulatoren z. B. in Mehrkanal- Fernsprechsystemen und für verschiedene Schaltaufgaben in Rechenmaschinen. Die Weltproduktion derartiger Dioden beträgt zur Zeit einige hundert Millionen Stück pro Jahr.

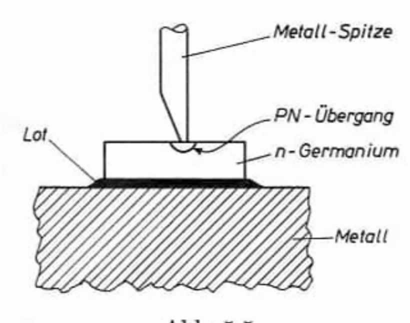


Abb. 5.5.
Aufbau einer Germanium-Punktkontaktdiode.

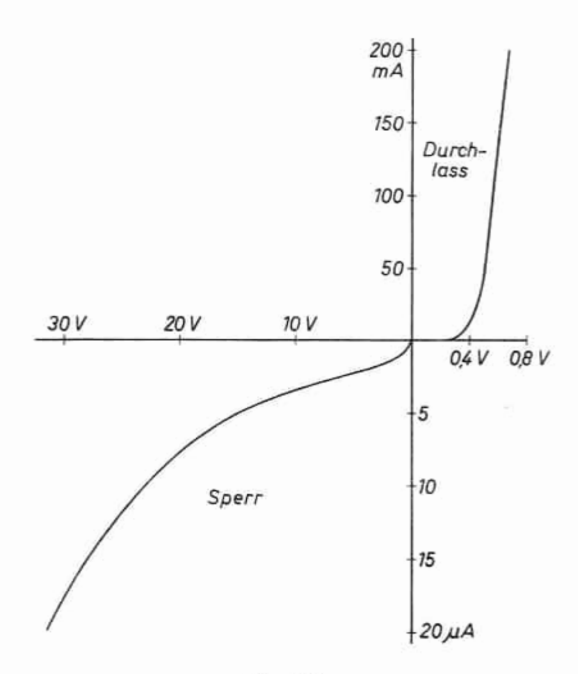


Abb. 5.6.
Kennlinie einer Germanium-Punktkontaktdiode
(z. B. gold bonded, Type AAY 13).

5.2.3 Zener-Diode. Die Sperrkennlinien geeignet konstruierter Siliziumdioden brechen bei ganz bestimmten kritischen Spannungen sehr steil ab (Abb. 5.7). Diese Erscheinung wird Zener-Effekt genannt und kann zur Spannungsstabilisierung verwendet werden, da der dynamische Widerstand jenseits des Abbruchs nur wenige Ohm beträgt, während er unterhalb der Abbruchspannung größer als 10 MΩ ist. Durch geeignete Kombination von drei Zener-Dioden kann eine Temperaturkonstanz von besser als $\pm$ 10^{-3} % der Spannung pro °C erreicht werden, was bereits den Anforderungen an ein Normalelement entspricht.

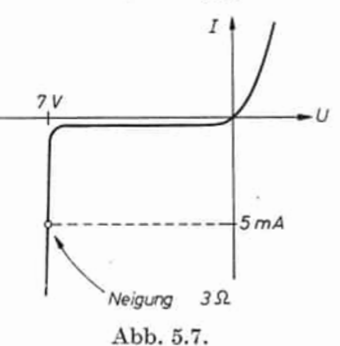


Abb. 5.7.
Kennlinie einer Zener-Diode (7-Volt-Type).

5.2.4 Kapazitätsdioden (*Varicap*). Auch diese Dioden enthalten einen einfachen *PN*-Übergang, bei dem aber jetzt die Tatsache ausgenützt wird, daß dieser *PN*-Übergang eine relativ große Kapazität darstellt, die gemäß Abb. 5.8 von der angelegten Spannung abhängt. Man kann dieses Bauelement z. B. zur automatischen Scharfabstimmung in Rundfunkempfängern einsetzen. Eine andere Anwendung ist die Verstärkung von Mikrowellen in sogenannten parametrischen Verstärkern.

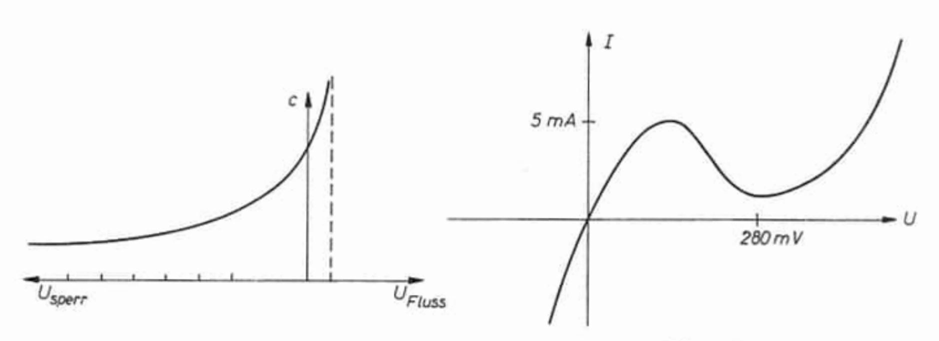


Abb. 5.8.
Kennlinie einer Kapazitätsdiode (Varicap).

Abb. 5.9.
Typische Kennlinie einer Tunneldiode.

5.2.5 Tunneldiode. Hier verwenden wir einen *PN*-Übergang, wo beide Seiten bis zur Entartung dotiert sind. Dadurch entsteht eine Kennlinie (Abb. 5.9), deren Flußrichtung ein Gebiet mit negativem differentiellem Widerstand enthält. Solche Dioden können zu Schaltzwecken in sehr schnellen Rechenmaschinen und wegen des negativen Widerstandes zur Verstärkung von höchsten Frequenzen verwendet werden.

5.2.6 Transistoren. Der Transistor ist zweifellos das bedeutendste Mitglied oder vielmehr die bedeutendste Gruppe von Mitgliedern der Familie der Halbleiterbauelemente. Die Weltproduktion wird für 1961 auf 300 Millionen Stück geschätzt. Die grundsätzliche Struktur eines Transistors ist in Abb. 5.10a

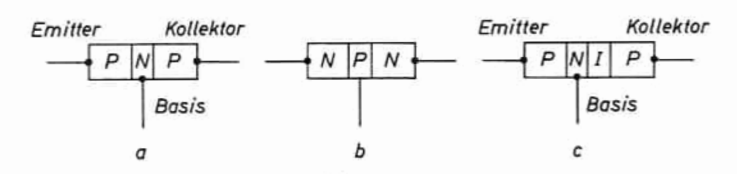
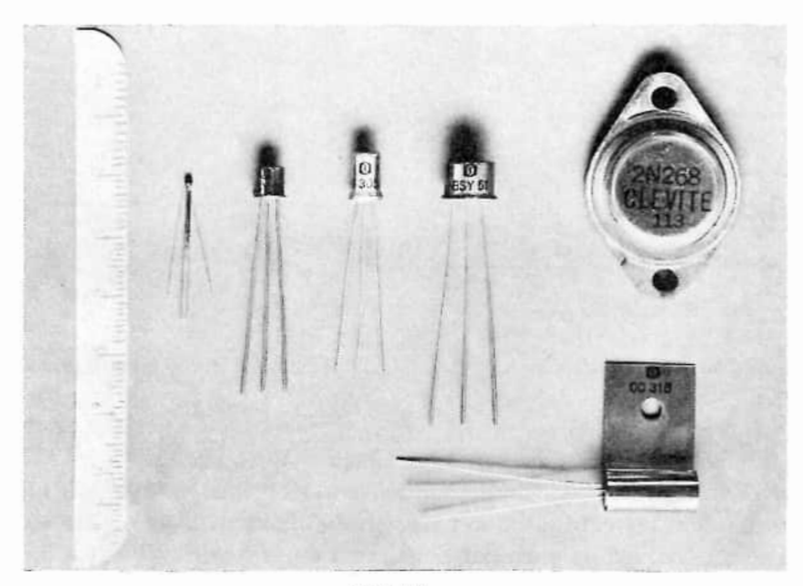


Abb. 5.10.
Schematischer Aufbau von Transistoren. a) PNP-Typ, b) NPN-Typ, c) Hochfrequenz-PNP-Typ mit Eigenleitungsschicht zwischen Basis und Kollektor.

und b gegeben. Sie besteht aus zwei *PN*-Übergängen, die eine dünne Zwischenschicht, die Basisschicht, gemeinsam haben, und drei sperrfreien Kontakten, dem Emitter-, dem Kollektor- und dem Basiskontakt. Der Verstärkereffekt dieser Halbleitertriode beruht auf der Wechselwirkung der beiden Ströme durch die beiden *PN*-Übergänge.

Die genaue Form und Größe dieser Transistoren richtet sich nach dem Anwendungszweck, d. h. danach, ob sie für hohe oder für niedrige Frequenz, große oder kleine Leistung, als lineare Verstärker oder als Schalter eingesetzt werden sollen. Danach richtet sich insbesondere auch die Herstellungsart, die



Tafel 8.
Verschiedene Transistortypen (Werkphoto Intermetall).

im nächsten Abschnitt besprochen werden soll. In Tafel 8 ist eine Auswahl verschiedener Transistortypen abgebildet. Abb. 5.11 gibt die Kennlinien eines Standardtransistors (OC 304) wieder. Hochfrequenztransistoren haben in der Regel die etwas kompliziertere Struktur der Abb. 5.10c, wo zwischen Basis und Kollektor noch eine eigenleitende Schicht (I) eingebaut ist [38]. Eine Variante dieser Struktur ist der Drift-Transistor [39].

5.2.7 Vierschichtdiode. Abb. 5.12 gibt die Struktur einer Vierschichtdiode wieder. Wir haben hier im ganzen sogar 3 PN-Übergänge, die über die engen Mittelzonen (5–15 μ) miteinander in Wechselwirkung treten [40, 41] und eine Kennlinie erzeugen, wie sie in Abb. 5.13 wiedergegeben ist. Diese Diode findet als typischer Schalter in Impulsgeneratoren, Ringzählern, Multivibratoren und ähnlichen Schaltungen Anwendung.

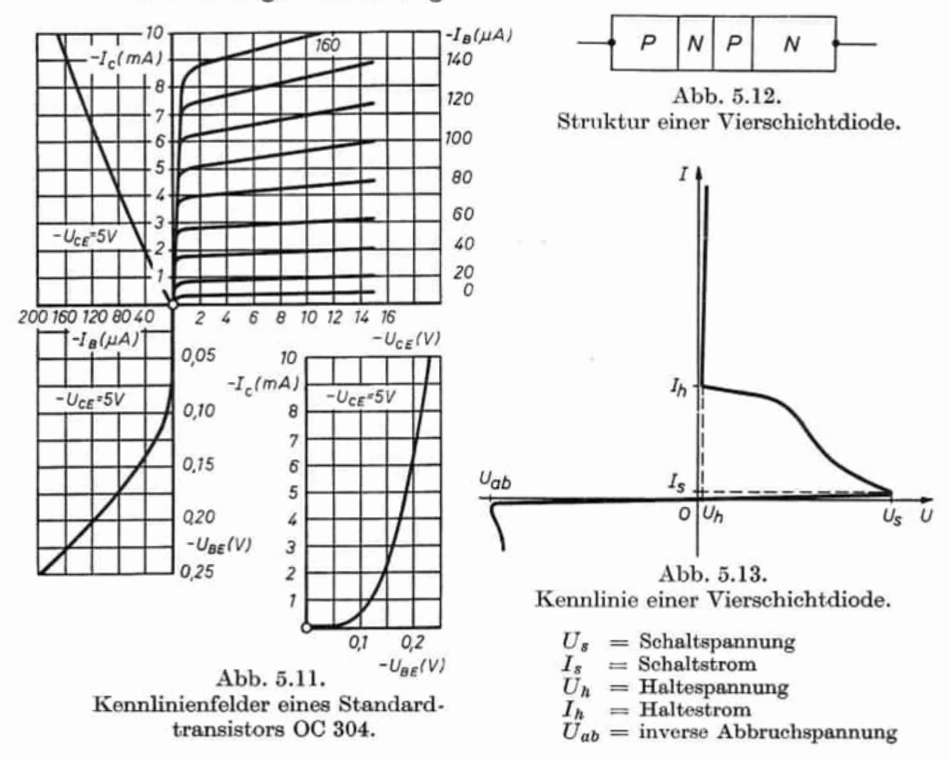


Abb. 5.11. Kennlinienfelder eines Standardtransistors OC 304.

Abb. 5.12. Struktur einer Vierschichtdiode.

Abb. 5.13. Kennlinie einer Vierschichtdiode.

U_s = Schaltspannung
I_s = Schaltstrom
U_h = Haltespannung
I_h = Haltestrom
U_{ab} = inverse Abbruchspannung

5.3 Herstellung von PN-Übergängen

Im letzten Abschnitt haben wir die Bedeutung des PN-Überganges für die Anwendung der Halbleiterphysik kennengelernt. Wir wollen uns jetzt mit der Technologie der Herstellung von PN-Übergängen befassen. Um saubere Verhältnisse zu erhalten und um insbesondere eine erhöhte Trägerrekombination am Übergang zu vermeiden, muß der PN-Übergang ganz im ungestörten, ein-

kristallinen Material liegen. Man kann also nicht einfach ein Stück p-Germanium und ein Stück n-Germanium in Kontakt bringen, sondern der PN-Übergang muß durch verschiedenes Dotieren der beiden Zonen im Einkristall selbst erzeugt werden. Das kann auf verschiedene Weise getan werden.

5.3.1 Gezogene PN-Übergänge. Bei dieser Methode wird die gewünschte Schichtenfolge bei der Herstellung des Kristalls erzeugt, indem der Kristall nach der CZOCHRALSKI-Methode aus der Schmelze gezogen wird (vgl. Abschnitt 2.2.2) und diese so dotiert ist, daß man zunächst ein hochohmiges Ende des Kristalls von einem Leitungstyp erhält. Während des Ziehens oder auch bei kurzem Anhalten des Ziehvorganges fügt man soviel von einem Störstoff entgegengesetzten Leitungstyps zur Schmelze zu, daß die gewünschte Trägerart in der gewünschten Konzentration in den Kristall eingebaut wird. Dabei ist zu beachten, daß die Löslichkeit der verschiedenen Störstoffe im festen Kristall verschieden ist und daß die wirksame Störstellendichte durch den Überschuß der miteinander konkurrierenden Störstoffe gegeben ist. Der in Abschnitt 5.2.6 genannte PNP-Transistorkristall kann also so hergestellt werden, daß man mit dem hochohmigen P-Teil durch Zugabe von In zur Schmelze beginnt, hernach gibt man soviel Sb zur Schmelze, daß die Wirkung des In aufgehoben wird und gleichzeitig die gewünschte überschüssige Donatorendichte erzeugt wird. Um eine dünne Mittelzone zu erhalten, gibt man bei nicht angehaltenem Ziehvorgang die dritte Dotierung kurz nach der zweiten in die Schmelze. Sie ist so bemessen, daß die Donatorendichte überkompensiert wird und die gewünschte Löcherleitfähigkeit des Emitters erreicht wird. Bei dieser Art der Herstellung läßt sich nur eine Folge von Schichten mit steigender Leitfähigkeit herstellen, weil geringe Konzentrationsvariationen beim Einbau der Störstoffe nicht ausgeschaltet werden können, so daß jeweils um ein Mehrfaches überdotiert werden muß. Zur Herstellung der einzelnen Elemente wird der Kristall parallel zur Zugrichtung in einzelne Blöckchen zersägt und die verschieden dotierten Teile mit nichtgleichrichtenden — sogenannten ohmschen — Kontakten versehen. In Abb. 5.14 sind eine Diode und ein NPN-Transistor dargestellt, die nach diesen Verfahren hergestellt sind. Da die P-Zone bei der Reinigungsätzung stärker angegriffen wird als das N-Material, entsteht im Falle des NPN-Transistors an der Oberfläche der Basisschicht eine Rinne, die das Auffinden dieser Schicht zum Zwecke des Kontaktierens erleichtert. Die Abb. 5.14b ist nicht maßstabgerecht gezeichnet. Die Basiszone ist in Wirklichkeit sehr dünn (ab 20 μ). Die ersten PNP- bzw. NPN-Transistoren sind tatsächlich nach diesem Verfahren hergestellt worden. Die Fabrikationsverfahren der beiden folgenden Abschnitte sind heute jedoch von größerer Bedeutung.

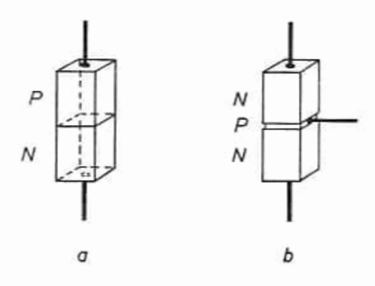


Abb. 5.14.
Gezogene PN-Übergänge.
a) Diode, b) NPN-Transistor.

5.3.2 Legierte PN-Übergänge. Das Legierverfahren besteht darin, daß auf einem Halbleiterplättchen vorbestimmter Größe mit Hilfe einer zweiten Behandlung die gewünschte Schichtenfolge erzeugt wird. Die Herstellung eines einfachen *PN*-Überganges kann beispielsweise so erfolgen, daß auf ein *N*-Ge-Plättchen gemäß Abb. 5.15 ein Stückchen In gebracht wird und Indium und Germanium auf eine definierte Temperatur erwärmt werden. Dabei löst das flüssig ge

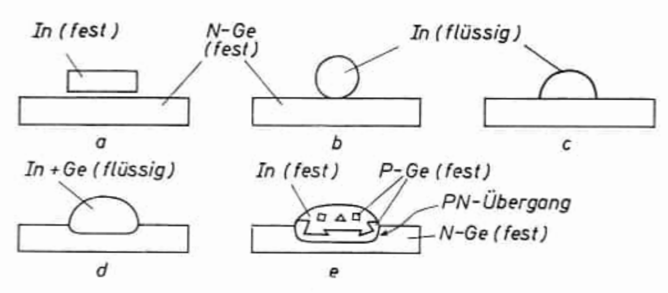


Abb. 5.15.
Indium-Germanium-Legierung.
a) Zimmertemperatur, b) Indium-Schmelzpunkt (155° C), c) Benetzung, d) Auflösung des Germaniums (400° C), e) nach dem Abkühlen.

wordene Indium so viel Germanium auf, bis die bei der gewählten Temperatur stabile Ge-Konzentration im In erreicht ist. Die bei der nachfolgenden Abkühlung anschließend an das nicht aufgelöste Ge sich bildende rekristallisierte Ge-Schicht enthält relativ viel In und ist deshalb *P*-leitend. Für die Gleichrichter- und Transistorwirkung sind, wie weiter oben angeführt, *PN*-Übergänge wünschenswert, wo die *P*-Schicht dieselbe einkristalline Struktur hat wie das nicht konvertierte *N*-Gebiet. Wenn der Rekristallisationsprozeß genügend sorgfältig durchgeführt wird, so kann eine relativ dicke Schicht als ein-

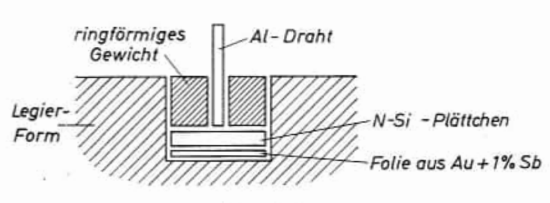
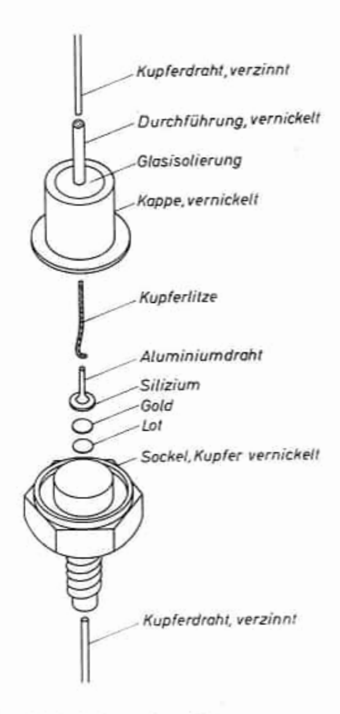
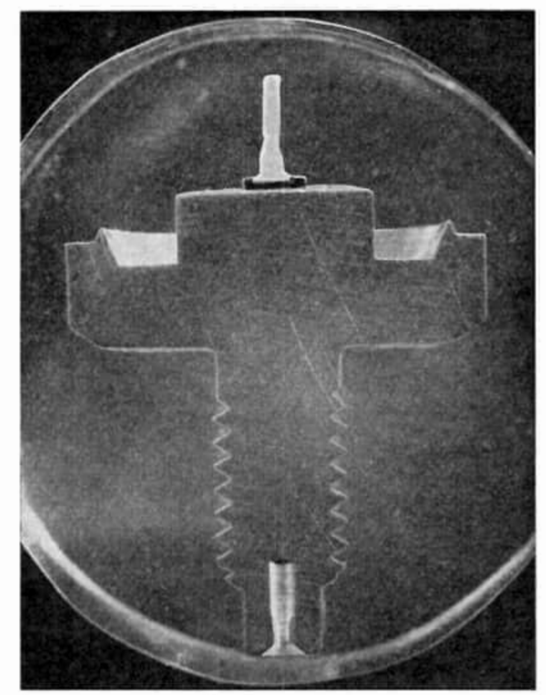


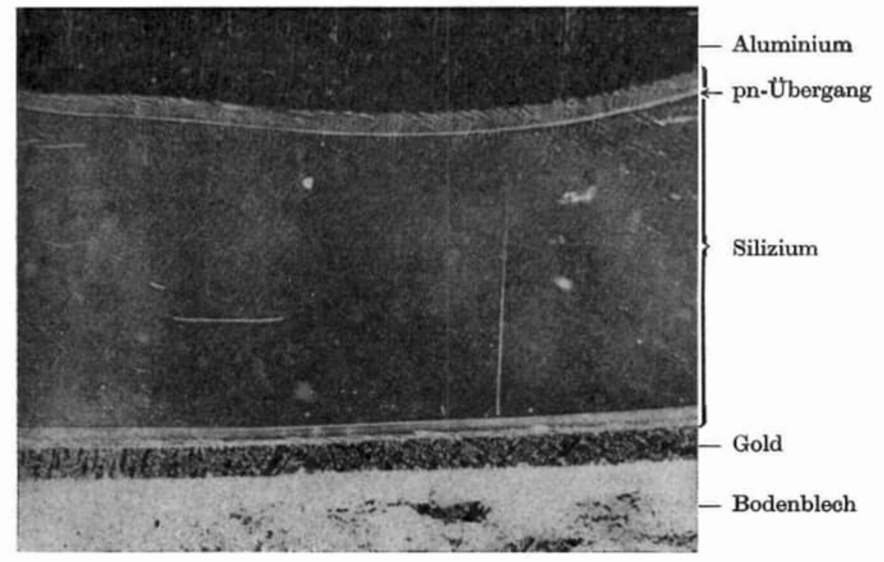
Abb. 5.16. a
Herstellung einer legierten Silizium-Diode.
Querschnitt durch die beladene Legierungsform.



b) Einbau ins Gehäuse.



Tafel 9. Querschnitt durch eine legierte Silizium-Diode (Werkphoto Intermetall).



Tafel 10. Vergrößertes Schliffbild einer legierten Silizium-Diode (nach ILSCHNER und STEINHÄUSER [95]).

kristalline Fortsetzung des N-Ge-Einkristalls entstehen. Die nicht mehr einkristallinen Gebiete stellen, zusammen mit dem schließlich festwerdenden In, einen sperrfreien und mechanisch spannungsfreien Kontakt für die P-Zone dar.

5.3.2.1 Beispiel: Legierte Silizium-Diode. Als erstes Beispiel für einen solchen Prozeß soll die Aluminium-Silizium-Legierung besprochen werden, wie sie z. B. zum Teil bei Fernsehgleichrichtern (Abschnitt 5.2.1 und Tafel 7) und ZENER-Dioden (Abschnitt 5.2.3) verwendet wird. Die Einzelteile werden dabei in Legierformen aus Sonderstahl oder Graphit zusammengehalten, in denen viele solcher Bauelemente in einem einzigen Ofendurchgang legiert werden. Einen Teil einer solchen Legierform, d. h. eine Einheit, ist in Abb. 5.16a vereinfacht im Querschnitt dargestellt. Den Einbau ins Gehäuse zeigt schematisch Abb. 5.16b. Die ganze Form wird in Wasserstoffatmosphäre im Durchlaufofen für wenige Minuten auf 650–700 °C erhitzt. Dabei legieren

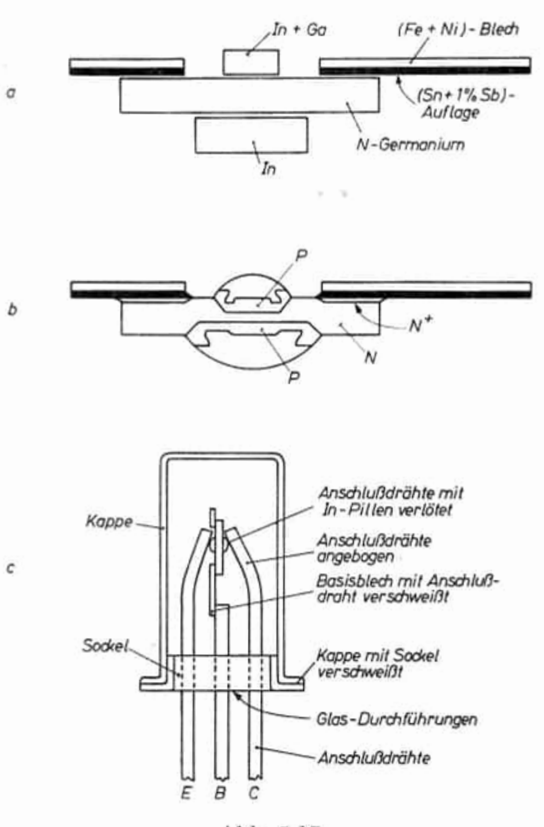
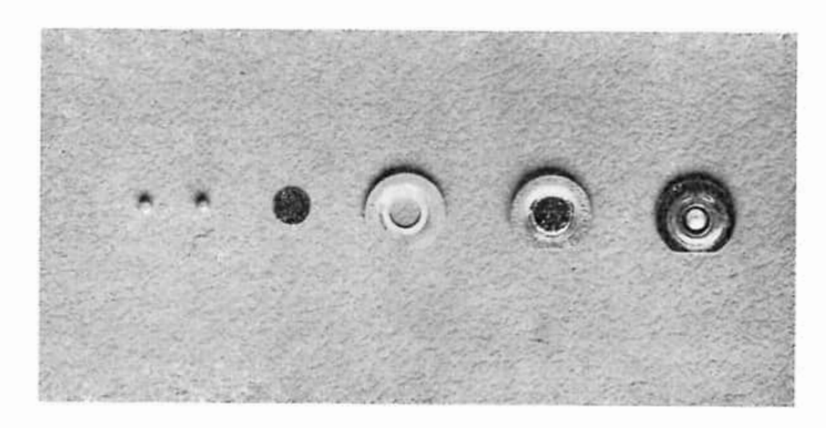
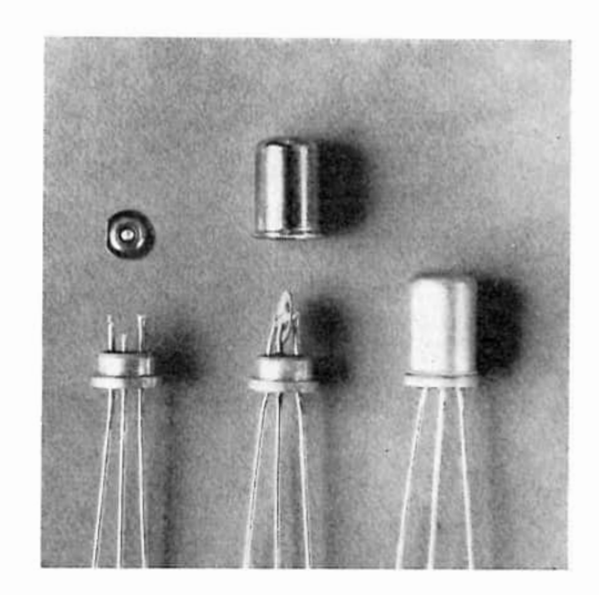


Abb. 5.17.
Herstellung eines legierten Germanium-Transistors
a) Anordnung vor dem Legieren, b) Querschnitt nach dem Legieren, c) Einbau ins Gehäuse.



Tafel 11.
Einzelteile eines Germanium-Transistors der Type OC 304 (Werkphoto Intermetall).



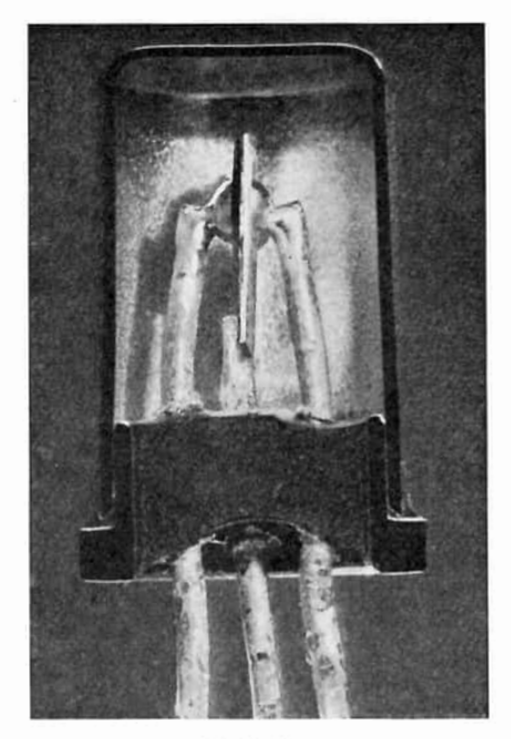
Tafel 12.
Montage und Verkappung einesGermanium-Transistors der Type OC 304 (Werkphoto Intermetall)

sowohl das untere Ende des Aluminiumdrahtes, als auch die mit Antimon versehene Goldfolie in das Siliziumplättchen ein, wobei die Legierverbindung mit dem Aluminium den *PN*-Übergang, diejenige mit dem Gold-Antimon den entarteten ohmschen Kontakt bildet. In Tafel 9 ist ein Querschnitt durch eine solche Diode dargestellt. Tafel 10 zeigt ein vergrößertes Schliffbild, auf dem alle diese verschiedenen Schichten sichtbar sind.

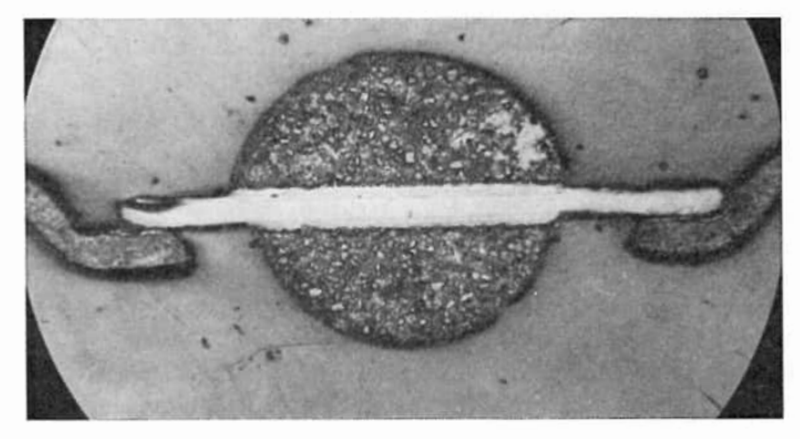
5.3.2.2 Beispiel: Legierter Germanium-Transistor. Nach dem in Abb. 5.15 dargestellten Verfahren werden heutzutage die meisten Germaniumtransistoren hergestellt. Kollektorpille, Germaniumkriställchen, Emitterpille und Basisblech werden gemäß Abb. 5.17 in geeigneten Legierformen aufeinandergelegt und in reduzierender Atmosphäre im Durchlaufofen zusammenlegiert. Das Basisblech soll einen großflächigen ohmschen Kontakt mit dem *N*-Germanium-Plättchen, d. h. mit der Basiszone herstellen. Zu diesem Zweck besteht es aus einem Nickel-Eisen-Gemisch, das denselben thermischen Ausdehnungskoeffizienten wie Germanium hat, und ist auf der einen Seite mit Zinn plus 1% Antimon überzogen. Bei der Legierung zwischen der Zinnschicht und dem Germanium sorgt der Antimongehalt dafür, daß die Rekristallisationsschicht stark *N*-Typ (N^+) und somit sperrfrei wird. Die Einzelteile für einen Transistor der Type OC 304 sind in Tafel 11 dargestellt. Der Einbau ins Gehäuse ist aus Abb. 5.17c, Tafel 12 und Tafel 13 zu ersehen. Ein vergrößertes Schliffbild zeigt Tafel 14.

5.3.2.3 Beispiel: Legierter Silizium-Transistor. Ganz entsprechend dem bisher Gesagten läuft auch die Herstellung des legierten Siliziumtransistors. Als Emitter und Kollektor können entweder Aluminiumdrähte wie in Abschnitt 5.3.2.1 oder Aluminiumpillen verwendet werden. Der Basiskontakt besteht z. B. aus einem Gold-Antimon-Ring. Tafel 15 zeigt einen Querschnitt durch einen Silizium-Aluminiumdraht-Transistor.

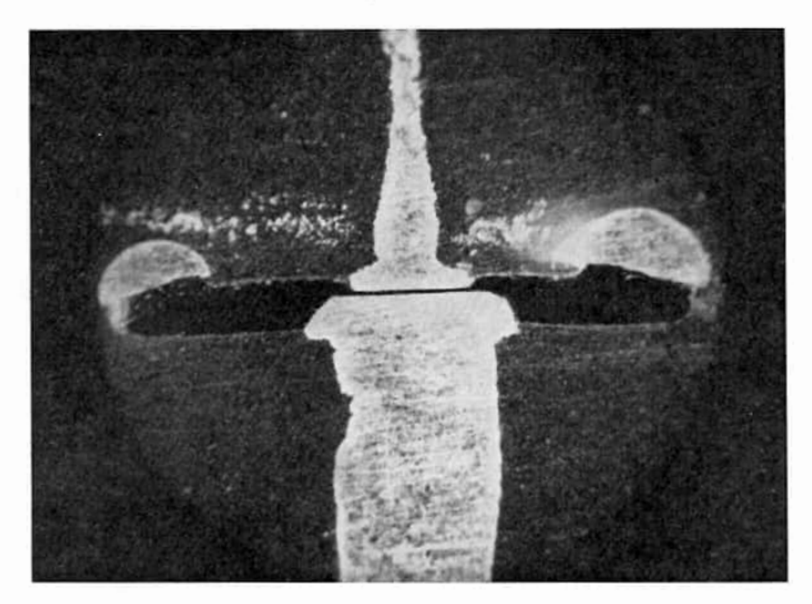
5.3.3 Diffundierte PN-Übergänge. Im dritten und modernsten Verfahren werden *PN*-Übergänge erzeugt, indem man die entsprechenden Störatome bei erhöhter Temperatur in den festen Halbleiterkristall eindiffundieren läßt, ohne daß ein Schmelzen des Kristalls oder eine flüssige Legierung auftritt. Die gewünschte Verunreinigung wird der Halbleiteroberfläche entweder als Gas in einem partiellen Vakuum angeboten oder in Form einer Verbindung auf die Oberfläche aufgebracht. Nach dem Erhitzen dieses Systems für längere Zeit (z. B. Stunden) auf Temperaturen bis zu 100–200 °C unterhalb des Schmelzpunktes entsteht eine Störstellenverteilung wie sie in Abb. 5.18 für die Diffusion eines Akzeptors in *N*-Material qualitativ dargestellt ist. Man sieht, daß dadurch eine *P*-Schicht an der Oberfläche entsteht. Die Dicke dieser Schicht, d. h. die Tiefe des *PN*-Überganges, hängt von der Oberflächenkonzentration, der Temperatur und der Zeit ab. Das sind Größen, die experimentell recht gut beherrscht werden können.



Tafel 13.
Querschnitt eines Germanium-Transistors der Type OC 304 (Werkphoto Intermetall).



Tafel 14.
Vergrößertes Schliffbild eines legierten Germanium-Transistors (Werkphoto Intermetall).



Tafel 15.
Vergrößertes Schliffbild eines legierten Silizium-Aluminiumdraht-Transistors (Werkphoto Intermetall).

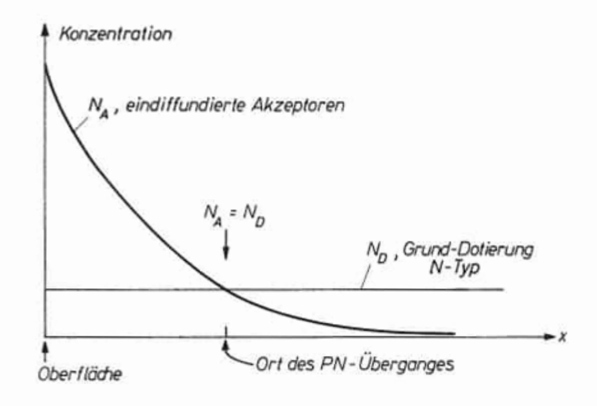


Abb. 5.18.
Konzentrationsprofil bei der Diffusion von Akzeptoren in N-Typ.

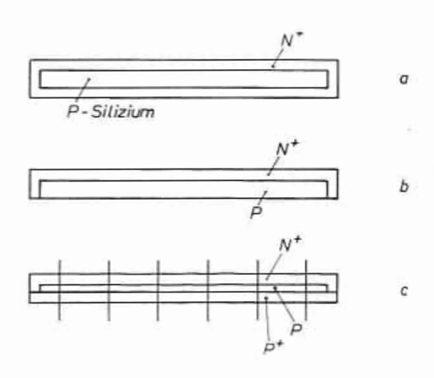


Abb. 5.19.
Herstellung eines diffundierten Silizium-Gleichrichters.

5.3.3.1 Beispiel: Diffundierter Silizium-Gleichrichter. Die wichtigsten Fabrikationsschritte sind aus Abb. 5.19 ersichtlich. Auf eine Scheibe aus P-Silizium wird Phosphorpentoxyd aufgebracht. Nach einigen Stunden im Ofen bei 1200–1300°C entsteht eine N-Typ-Schicht von etwa 50μ Dicke (Abb. 5.19a). Diese Schicht wird dann einseitig durch Läppen beseitigt (Abb. 5.19b); es wird Bortrioxyd aufgebracht und wieder diffundiert. Man erhält auf diese Weise die Schichtenfolge N^+–P–P^+. Nach der Diffusion wird die Scheibe beiderseits platiert und in Plättchen geeigneter Größe zersägt (Abb. 5.19c), deren jedes nach dem Einbau in ein Gehäuse einen Gleichrichter ergibt.

5.3.3.2 Germanium-Mesatransistor. Der Germanium-Mesatransistor ist derjenige Transistortyp, mit dem die bisher höchsten Frequenzen von allen

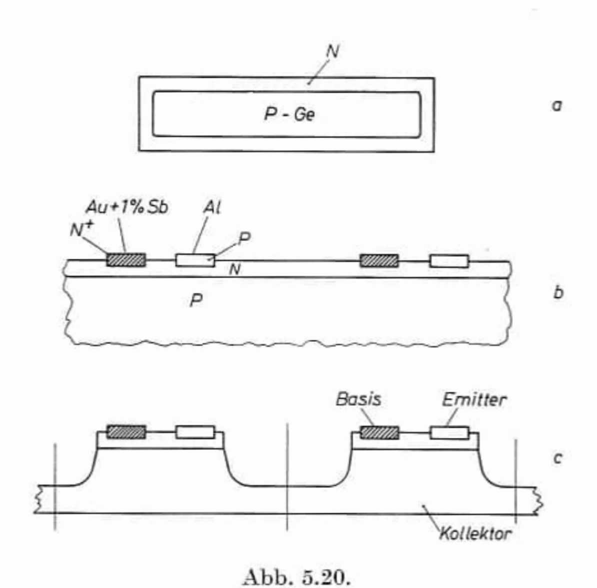
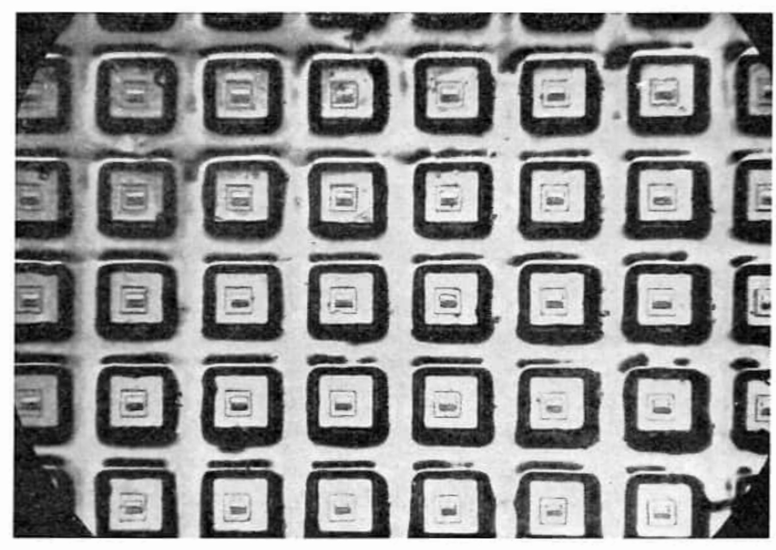


Abb. 5.20.
Herstellung eines Germanium-Mesatransistors.
a) Diffusion der Basisschicht. b) Aufdampfen und Einlegieren der Emitter- und Basisflecken. c) Mesaätzung, Abtragung der Rückseite, Trennung in Einzelelemente.



Tafel 16.
Germaniumplatte mit fertigen Mesa-Strukturen
(Werkphoto Intermetall).

Transistortypen erreicht wurden (mehr als 1 GHz). Der Kollektorübergang wird durch eine etwa $5\,\mu$ tiefe Diffusion von Antimon in *P*-Germanium hergestellt (Abb. 5.20a). Wegen der Flachheit der Diffusion entsteht eine *NIP*-Struktur, die derjenigen des Kollektors von Abb. 5.10c ähnelt. Es werden dann auf eine Seite der Germaniumscheibe in regelmäßigen Abständen Paare von Metallflecken aufgedampft und etwa $3\,\mu$ tief einlegiert. Von diesen Fleckenpaaren ist der eine ein Aluminiumfleck, der andere ein Fleck aus Gold mit 1% Antimon (Abb. 5.20b). Dadurch entstehen auf der *N*-Schicht ein *PN*-Übergang und ein Basisanschluß. Nun maskiert man die unmittelbare Umgebung mit einem säurefesten Wachs und ätzt die ganze Scheibe, so daß an den Orten der Flecken kleine Erhöhungen übrigbleiben, sogenannte „Mesa"-Berge (von dem spanischen Wort für Tafelberg) (siehe Tafel 16). DieRückseite wird abgeschliffen oder abgeätzt und dann die Scheibe in die einzelnen Elemente zerteilt (Abb. 5.20c), die dann an der Unterseite (Kollektor) und an den beiden Metallflecken (Emitter und Basis) kontaktiert und in Gehäuse eingebaut werden.

5.4 Gleichstromverhalten eines PN-Überganges

5.4.1 Allgemeines. Als einfachste *PN*-Schicht denken wir uns die beiden Hälften eines Ge-Einkristalls verschieden stark, aber räumlich konstant, mit Akzeptoren bzw. Donatoren versehen. Über eine Abszisse x, die wir senkrecht zur *PN*-Grenze

durch den Kristall legen, sind in Abb. 5.21 Störstellen- und Stromträgerdichte aufgetragen, und zwar zunächst für den Fall, daß die beiden Kristallhälften noch nicht im Austausch ihrer beweglichen Träger stehen. Da wir Dichten sehr verschiedener Größe darzustellen haben, ist für die Ordinate ein logarithmischer Maßstab gewählt worden. Entsprechend unserer Voraussetzung haben wir links eine konstante Akzeptordichte N_A. Ihr entspricht bei Zimmertemperatur nach Abschnitt 2.4.2.3 eine ebenso große Löcherdichte $p_P = N_A$. Nach Gl. 4.6 sind in der linken Kristallhälfte außer den Löchern auch Elektronen vorhanden, und zwar in einer solchen Dichte n_P, daß $p_P \cdot n_P = n_i^2$ ist. Ebenso entspricht

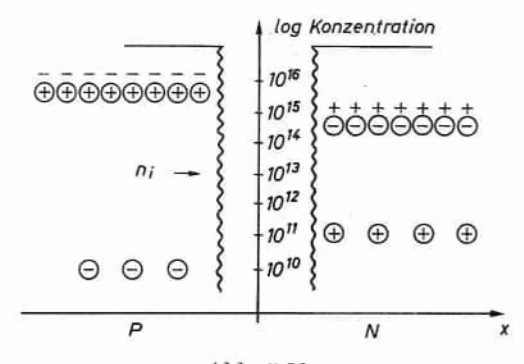


Abb. 5.21.
Störstellen- und Trägerdichten zu beiden Seiten eines PN-Übergangs. (Die Raumladungszone ist nicht mitgezeichnet.)

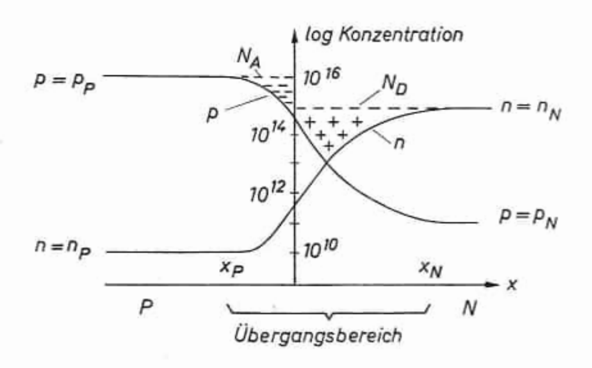


Abb. 5.22.
Trägerdichten im PN-Übergang bei Gleichgewicht.

der Donatorendichte N_D in der rechten Kristallhälfte eine ebenso große Elektronendichte n_N. Da in unserem Beispiel n_N kleiner als p_P ist, ist die Löcherdichte p_N rechts größer als die Elektronendichte n_P links.

Um anzudeuten, daß die Stromträger einem elektrischen Feld folgen können, werden in Abb. 5.21 und später Elektronen und Löcher mit einem Ring versehen. Die geladenen Störstellenionen sind unbeweglich und deshalb ohne Ring gezeichnet.

Für alle Temperaturen gilt nach Gl. 2.18 bzw. Gl. 4.34

$$pn = n_i^2 \sim e^{-E_G/kT} \qquad \text{(Gl. 5.1)}$$

Da im weiten Bereich um Zimmertemperatur alle Störstellen ionisiert sind, ändert sich die Majoritätendichte der Stromträger mit der Temperatur fast nicht, dafür fällt die gesamte Temperaturabhängigkeit gemäß Gl. 4.16 und Gl. 5.1 auf die Minoritätendichten.

Bringt man die beiden Kristallhälften von Abb. 5.21 zusammen, so findet wegen ihrer verschiedenen Konzentration auf beiden Seiten ein Ausgleich der beweglichen Stromträger durch Diffusion statt: Defektelektronen der linken Kristallhälfte fließen nach rechts, Elektronen von der rechten in die linke Hälfte. Bei diesem Trägerausgleich bleiben die geladenen Störstellen stehen und bilden eine Raumladung, die den Stromträgerausgleich auf einen verhältnismäßig engen Bereich begrenzt. Da in diesem Bereich ein elektrisches Feld vorhanden ist, nennen wir ihn Feldzone oder auch Raumladungszone. Oft spricht man auch von einem Übergangsbereich, weil in dieser Zone der Ausgleich der Stromträger stattfindet. Die Konzentrationen, die sich im Übergangsgebiet auf Grund dieses Ausgleichs einstellen, sind in Abb. 5.22 dargestellt.

Wird berücksichtigt, daß die Konzentrationen p und n logarithmisch aufgetragen sind, so erkennt man, daß die P-Seite der Übergangszone weitgehend entblößt ist von Löchern und die N-Seite von Elektronen. Wird an den PN-Übergang eine Spannung angelegt, so daß ein zusätzliches elektrisches Feld in der Übergangszone auftritt, das von der P- nach der N-Seite zeigt (also von links nach rechts in Abb. 5.22), so verschiebt sich die Verteilung der Majoritäten so, daß sich die Raumladungszone verengt. D. h. die Löcher werden nach rechts, die Elektronen nach links gedrückt. Beim Anlegen einer Spannung in der umgekehrten Richtung verbreitert sich die Raumladungszone. Das Anlegen einer Wechselspannung führt demnach zu einem fortgesetzten Laden und Entladen der beiden Randgebiete an beiden Seiten der Raumladungszone. Dies kann durch eine Kapazität beschrieben werden, und zwar hängt diese Kapazität von der Dicke der Raumladungszone ab. Diese wiederum ist, wie wir oben sahen, eine Funktion der angelegten Gleichspannung. Wir verstehen also jetzt bereits in qualitativer Weise die Spannungsabhängigkeit der Kapazität des PN-Übergangs, d. h. die Wirkungsweise der Kapazitätsdiode von Abschnitt 5.2.4 und Abb. 5.8.

Bei einem zusätzlichen Feld von links nach rechts werden aber auch Löcher vom P-Gebiet ins N-Gebiet hinübergetrieben und Elektronen vom N- ins P-Gebiet, wodurch ein Strom zustande kommt. Da im P-Gebiet viel Löcher und im N-Gebiet viel Elektronen vorhanden sind, wird dieser Strom groß sein (Flußrichtung z. B. in Abb. 5.4).

Bei einem zusätzlichen Feld von rechts nach links kann ein Strom nur so zustande kommen, daß Elektronen aus dem P-Gebiet ins N-Gebiet gezogen werden und Löcher aus dem N- ins P-Gebiet. Die Elektronen im P-Gebiet und die Löcher im N-Gebiet sind aber nur in sehr kleiner Konzentration vorhanden, und die Nachlieferung ist durch die spontane, thermische Generation dieser Minoritätsträger begrenzt. Es fließt deshalb nur ein kleiner Strom (Sperrrichtung), der in weitem Bereich unabhängig von der angelegten Spannung ist. Erst bei sehr hohen Sperrspannungen treten Durchbrucherscheinungen, z. B. der sogenannte ZENER-Effekt, ein (Abb. 5.7).

Diese qualitativen Überlegungen sollen in den folgenden Abschnitten vertieft und quantitativ unterbaut werden.

5.4.2 Feldzone

5.4.2.1 Feldzone ohne äußere Spannung. Im folgenden soll der Verlauf der Trägerdichte und des elektrischen Potentials in der Feldzone und ihre Breite für die oben angenommenen Verhältnisse diskutiert werden. Nach den Ausführungen in Abschnitt 4.1.4 läßt sich die Trägerdichte an jeder Stelle angeben, wenn das elektrische Potential der Doppelschicht, die durch die Ladungsverschiebung entstanden ist, an jeder Stelle bekannt ist. Anschaulich ist sofort, z. B. aus Abb. 5.22, zu sehen, daß in der Doppelschicht ein Feld von der positiven Raumladung der entblößten Donatoren zur negativen Raumladung der entblößten Akzeptoren, also von rechts nach links zeigend, vorhanden sein muß. Demnach muß das elektrische Potential φ (gemäß Gl. 4.42) von links nach rechts ansteigen, was in Abb. 5.23 gezeichnet ist. Die Gesamtzunahme sei $\varphi(x_N) - \varphi(x_P) = V_D$.

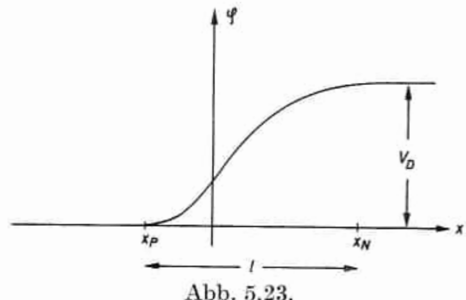


Abb. 5.23.
Potentialverlauf im PN-Übergang bei Gleichgewicht.

Nach dem BOLTZMANNschen Satz (Gl. 4.43 und 4.44) ist

$$p_N = p_P\, e^{\frac{-\varphi(x_N) + \varphi(x_P)}{kT/e}} = p_P\, e^{-V_D/U_T} \qquad \text{(Gl. 5.2)}$$

und

$$n_P = n_N\, e^{\frac{\varphi(x_P) - \varphi(x_N)}{kT/e}} = n_N\, e^{-V_D/U_T} \qquad \text{(Gl. 5.3)}$$

Demnach müßte $p_N/p_P = n_P/n_N$ sein, was wegen Gl. 5.1 der Fall ist.

Man erhält also die Potentialänderung V_D

$$V_D = U_T \ln \frac{p_P}{p_N} = U_T \ln \frac{n_N}{n_P} = U_T \ln \frac{p_P\, n_N}{n_i^2} \qquad \text{(Gl. 5.4)}$$

Die Größe V_D wird Diffusionsspannung genannt, weil sie durch die Diffusion der Majoritäten in das Gebiet entgegengesetzten Leitungstyps aufgebaut wird. Sie beträgt in der Regel einige Zehntel Volt, für die Konzentrationen der Abb. 5.22 ergibt sich z. B. im Fall des Germaniums $V_D = 0{,}27$ Volt.

Allgemein ist nach Abschnitt 4.1.4, wenn wir setzen,

$$p(x) = p_P\, e^{-\varphi(x)/U_T} \qquad \text{(Gl. 5.5)}$$

und

$$n(x) = n_P\, e^{+\varphi(x)/U_T} = n_N\, e^{(\varphi(x) - V_D)/U_T} \qquad \text{(Gl. 5.6)}$$

Daraus geht hervor, daß die Kurven von Abb. 5.22 und 5.23 einander ähnlich sein müssen, weil die Konzentrationen p und n in Abb. 5.22 in logarithmischem Maßstab aufgetragen worden sind.

Wir wollen jetzt die genaue Form dieser Kurven, d. h. die Funktion $\varphi(x)$ quantitativ berechnen. Dazu steht die POISSON-Gleichung (Gl. 4.63) zur Verfügung. Die darin vorkommende Ladungsdichte ist

$$\varrho = e\,(N_D - N_A + p - n) \qquad \text{(Gl. 5.7)}$$

Für die nun folgende Rechnung wollen wir uns auf den Fall unseres einfachen, abrupten PN-Übergangs beschränken, für den im N-Gebiet der Raumladungszone $N_A = 0$ ist. Außerdem ist dort $p \ll N_D$. Wegen des exponentiellen Zusammenhangs zwischen n und φ (Gl. 5.6) kann n gegenüber N_D schon vernachlässigt werden, wenn φ nur wenige Einheiten U_T unter V_D liegt. Dies ist schon bei Werten von x der Fall, die nur wenig kleiner als x_N sind. Deshalb gilt $n \ll N_D$ als gute Näherung für fast die ganze N-Hälfte der Raumladungszone

$0 < x < x_N$. Somit kann für dieses Gebiet die Gl. 5.7 sehr vereinfacht werden, und wir schreiben die POISSON-Gleichung für unseren eindimensionalen Fall:

$$\frac{d^2\varphi}{dx^2} = -\frac{e}{\varepsilon\,\varepsilon_0} N_D \qquad \text{(Gl. 5.8)}$$

Entsprechend ist für das P-Gebiet $N_D = 0$ und $p, n \ll N_A$. Und deshalb

$$\frac{d^2\varphi}{dx^2} = \frac{e}{\varepsilon\,\varepsilon_0} N_A \qquad \text{(Gl. 5.9)}$$

Diese Gleichungen lassen sich bequem integrieren. Für das N-Gebiet erhalten wir aus Gl. 5.8 mit der Randbedingung

$$\left.\frac{d\varphi}{dx}\right|_{x = x_N} = 0 \qquad \text{(Gl. 5.10)}$$

die Gleichung

$$\frac{d\varphi}{dx} = \frac{e}{\varepsilon\varepsilon_0} N_D (x_N - x) \qquad \text{(Gl. 5.11)}$$

Und nach der zweiten Integration mit der Randbedingung

$$\varphi(x_N) = V_D \qquad \text{(Gl. 5.12)}$$

erhält man für die N-Seite

$$\varphi = V_D - \frac{e}{2\,\varepsilon\,\varepsilon_0} N_D (x_N - x)^2 \qquad \text{(Gl. 5.13)}$$

Entsprechend findet man durch zweifache Integration von Gl. 5.9 mit

$$\left.\frac{d\varphi}{dx}\right|_{x = x_P} = 0 \quad \text{und} \quad \varphi(x_P) = 0$$

das Potential für die P-Seite

$$\varphi = \frac{e}{2\varepsilon\,\varepsilon_0} N_A (x - x_P)^2 \qquad \text{(Gl. 5.14)}$$

Wie man sieht, ist die Kurve der Abb. 5.23 zusammengesetzt aus zwei Parabeln, deren Scheitel bei x_P und x_N liegen und deren Krümmungen durch N_A bzw. N_D bestimmt sind.

Die Größen x_P und x_N, d. h. die Ausdehnung der Raumladungszone, können aus dem stetigen Anschluß von φ und $\frac{d\varphi}{dx}$ der beiden Teillösungen in Punkt $x = 0$ gewonnen werden. So ergibt die Gleichheit von $\frac{d\varphi}{dx}$ aus Gl. 5.13 und 5.14 für $x = 0$:

$$N_D\, x_N = -N_A\, x_P = N_A\, |x_P| \qquad \text{(Gl. 5.15)}$$

Dies bedeutet, daß die positive Ladung aller entblößten Donatoren gleich der negativen Ladung aller entblößten Akzeptoren sein muß; d. h. im ganzen ist die *PN*-Schicht neutral. Gl. 5.15 zeigt auch, daß bei einem symmetrischen Übergang ($N_D = N_A$) die Raumladungszone sich symmetrisch in beide Zonen erstrecken wird ($x_N = |x_P|$), während für den unsymmetrischen Fall die Raumladungszone sich mehr in das schwächer dotierte, d. h. höherohmige Gebiet erstrecken wird. Z. B. ergibt sich für $N_A \gg N_D$, $|x_P| \ll x_N$. Die Betrachtung der Gl. 5.13 und 5.14 führt auf

$$\varphi(x_N) - \varphi(0) = \frac{e}{2\varepsilon\varepsilon_0} N_D x_N^2 \qquad \text{(Gl. 5.16)}$$

und

$$\varphi(0) - \varphi(x_P) = \frac{e}{2\varepsilon\varepsilon_0} N_A x_P^2 \qquad \text{(Gl. 5.17)}$$

Daraus ergibt sich unter Berücksichtigung von Gl. 5.15, daß beim unsymmetrischen *PN*-Übergang auch der größere Teil der Spannung im hochohmigen Teil abfällt. Dies ist unmittelbar anschaulich aus Abb. 5.23 zu sehen.

Der Anschluß von φ bei $x = 0$ ergibt

$$V_D - \frac{e}{2\varepsilon\varepsilon_0} N_D x_N^2 = \frac{e}{2\varepsilon\varepsilon_0} N_A x_P^2 \qquad \text{(Gl. 5.18)}$$

Durch Kombination von Gl. 5.15 mit 5.18 erhält man schließlich

$$x_N = \frac{N_A}{N_D + N_A} l \qquad \text{(Gl. 5.19)}$$

$$x_P = \frac{N_D}{N_D + N_A} l \qquad \text{(Gl. 5.20)}$$

wobei l die gesamte Dicke der Raumladungszone ist

$$l \equiv x_N - x_P = \left(\frac{2\varepsilon\varepsilon_0}{e} \frac{N_D + N_A}{N_D N_A} V_D\right)^{\frac{1}{2}} \qquad \text{(Gl. 5.21)}$$

Diese Dicke der Feldzone ist im allgemeinen von der Größenordnung μ. Für die in Abb. 5.22 eingezeichneten Konzentrationen erhält man z. B.

$$\begin{aligned} l &= 0{,}76\,\mu \\ x_N &= 0{,}69\,\mu \\ x_P &= 0{,}069\,\mu \end{aligned}$$

Die obige Rechnung gilt nur für den abrupten Übergang. Liegt z. B. ein linearer Störstellengradient

$$b \equiv \frac{d\,(N_D - N_A)}{dx}$$

vor, so ist an Stelle von Gl. 5.7 zu schreiben [33]

$$\varrho = e(bx + p - n) \approx ebx \qquad \text{(Gl. 5.7a)}$$

und statt Gl. 5.8 und 5.9

$$\frac{d^2\varphi}{dx^2} = -\frac{e}{\varepsilon\varepsilon_0}bx \qquad (x_P < x < x_N) \qquad \text{(Gl. 5.8a)}$$

In diesem Fall ist das Problem symmetrisch, also $x_N = -x_P$. Es gelten wieder die Randbedingungen Gl. 5.10, Gl. 5.12 und $\varphi(x_P) = 0$. Dann erhält man nach zweimaliger Integration an Stelle von Gl. 5.21

$$l = 2\,x_N = \left(\frac{12\,\varepsilon\varepsilon_0 V_D}{e\,b}\right)^{\frac{1}{3}} \qquad \text{(Gl. 5.21a)}$$

Auf diese Weise kann im Prinzip jede beliebige Störstellenverteilung im *PN*-Übergang behandelt werden. In der Praxis kommt man jedoch meistens mit den beiden oben beschriebenen Fällen aus. So wird z. B. ein legierter *PN*-Übergang in der Regel recht gut durch Gl. 5.21 wiedergegeben, während ein diffundierter *PN*-Übergang durch den linearen Fall von Gl. 5.21a angenähert werden kann.

Mancher Leser wird überrascht sein, daß innerhalb eines Halbleiters, an dessen Enden keine äußere Spannung angelegt wurde, die also auf gleichem Potential (z. B. Erdpotential) sind, elektrische Felder und damit von Null verschiedene „Potentiale" vorliegen. Das rührt daher, daß man die Potentiale, die beim Kontakt zweier verschiedener elektrisch leitender Stoffe entstehen (Kontaktpotentiale oder Voltaspannung), oft unterschlägt. Grundsätzlich entsteht immer ein Kontaktpotential (besser Kontaktpotentialunterschied), wenn man zwei verschiedene Stoffe miteinander verbindet. Z. B. entsteht es beim Erden eines Ni-Drahtes an der Cu-Erde und hat an beiden Verbindungsstellen dieselbe Größe (Abb. 5.24). Oft wird es nicht bemerkt, weil die Elektronen die Energie, die sie zur Überwindung der einen Schwelle benötigen, beim Übertritt über die andere Schwelle wieder abgeben. Grundsätzlich ist ein Kontaktpotential jedoch nachzuweisen.

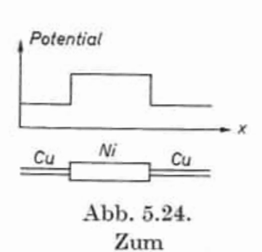


Abb. 5.24. Zum Kontaktpotential.

N- und *P*-Ge sind zwei verschiedene Stoffe; sie haben z. B. sehr verschiedene Elektronendichten, und bei ihrem Kontakt entsteht deshalb ein Kontaktpotential. Wenn beide Enden eines *PN*-Überganges beispielsweise mit einem Cu-Draht geerdet werden, so sind die Kontaktpotentiale an den verschiedenen Enden ungleich, so daß die Summe aller Kontaktpotentiale verschwindet. Für den weiter interessierten Leser sei auf das Buch von E. Spenke, Kapitel X, verwiesen.

5.4.2.2 Darstellung des PN-Übergangs im Bändermodell. Bevor wir mit der Behandlung der physikalischen Vorgänge im PN-Übergang fortfahren, soll kurz skizziert werden, wie die Potentiale und Konzentrationen im PN-Übergang mit Hilfe des in Abschnitt 2.4.3 eingeführten Bändermodells dargestellt werden können. Diese Darstellung fügt zwar zu der bisher gewählten Darstellungsweise für den Fall des nichtentarteten PN-Übergangs physikalisch nichts hinzu und kann deshalb beim ersten Durchlesen übergangen werden. Sie ist jedoch bei der Betrachtung von komplizierteren Problemen wegen ihrer Übersichtlichkeit und Eleganz ein wichtiges Hilfsmittel. Insbesondere aber ist sie auch dann noch brauchbar, wenn es sich um PN-Übergänge im entarteten Material handelt, wo die Boltzmann-Statistik mit den Gleichungen 5.1–5.6 durch die Fermi-Statistik ersetzt werden muß. Wir werden davon in Abschnitt 5.7 Gebrauch machen.

In Abb. 2.21 ist die Energie E eines Elektrons als Funktion des Ortes x aufgetragen. Da im homogenen Material bei Gleichgewicht die Energie E von x unabhängig ist, sind die verschiedenen Niveaus, z. B. der obere und untere Bandrand, dort als horizontale Linien gezeichnet. Das ist im allgemeinen, insbesondere im PN-Übergang, nicht mehr der Fall. Zu der in Abb. 2.21 gezeichneten Energie E ist die elektrostatische Energie auf Grund des elektrischen Potentials von Abb. 5.23, also $-e\varphi$, hinzuzuzählen. Dies ist in Abb. 5.25 geschehen. Dabei ist der Nullpunkt von E so gewählt worden, daß die gestrichelte Linie in Abb. 5.25 den negativen Wert des in Abb. 5.23 definierten Potentials φ darstellt (bis auf den Faktor e). Diese gestrichelte Linie liegt etwa in der Mitte der Bandlücke, genaugenommen um den Betrag $1/2\ kT\ ln\ (N_C/N_V)$ unterhalb der Mitte.

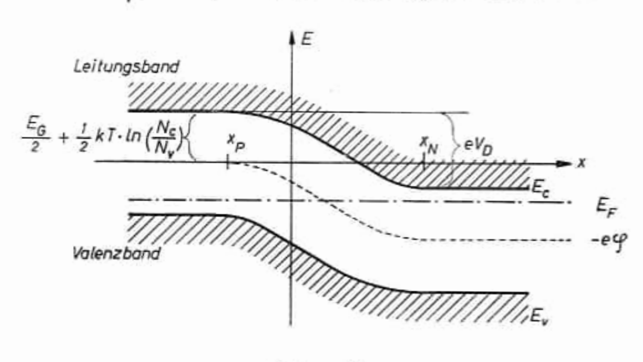


Abb. 5.25.
Bändermodelldarstellung des PN-Übergangs.

Die strichpunktierte Gerade ist die Fermikante E_F. Sie ist für den vorliegenden Fall des thermischen Gleichgewichts unabhängig von x, wie schon in Abschnitt 4.1.4 ohne Beweis erwähnt wurde. Die Gerade E_F liegt deshalb in Abb. 5.25 horizontal. Ihre vertikale Position ist durch n_N nach Gl. 4.30 gegeben:

$$n_N = N_C e^{-(E_C(x_N) \qquad)/kT} \qquad \text{(Gl. 5.22)}$$

Andererseits ist für die ungestörte P-Seite mit $p = p_P$ auch Gl. 4.32 erfüllt

$$p_P = N_V e^{-(E_F - E_V(x_P))/kT} \qquad \text{(Gl. 5.23)}$$

Die Multiplikation dieser beiden Gleichungen ergibt wegen der Konstanz von E_F unter Berücksichtigung von Gl. 4.34 und der Definition

$$E_C(x_P) - E_C(x_N) = eV_D \qquad \text{(Gl. 5.24)}$$

die Gleichung

$$n_N\, p_P = N_C\, N_V\, e^{\frac{E_C(x_P) - E_V(x_P) - eV_D}{kT}} = n_i^2 e^{V_D / U_T} \qquad \text{(Gl. 5.25)}$$

die mit Gl. 5.4 identisch ist. Mit den aus Abb. 5.25 abzulesenden Bezeichnungen

$$E_C = E_C(x_N) + e\,V_D - e\varphi(x) \qquad \text{(Gl. 5.26)}$$

$$E_V = E_V(x_P) - e\varphi(x)$$

ergeben sich aus Gl. 4.30 und 4.32 auch zwanglos die Gleichungen 5.5 und 5.6. Man sieht also, daß unsere Abb. 5.25 zusammen mit den Gleichungen der vereinfachten FERMI-Statistik Gl. 4.30 und 4.32 schon alles enthält.

Wir können die Konzentrationen der Elektronen und Löcher auch sehr einfach aus Abb. 5.25 ablesen, wenn wir unsere Nullpunktfestsetzung beachten,

wonach
$$E_C = \frac{E_G}{2} + kT \ln \sqrt{\frac{N_C}{N_V}} - e\,\varphi \qquad \text{(Gl. 5.27)}$$

Dies ergibt mit Gl. 4.30 und 4.34

$$n = N_C\, e^{-\frac{E_C - E_F}{kT}} = \sqrt{N_C N_V\, e^{-\frac{E_G}{2kT}}}\; e^{\frac{E_F + e\varphi}{kT}} = n_i e^{\frac{E_F - (-e\varphi)}{kT}} \qquad \text{(Gl. 5.28)}$$

und entsprechend
$$p = n_i\, e^{\frac{(-e\varphi) - E_F}{kT}} \qquad \text{(Gl. 5.29)}$$

Die Größen n und p sind demnach durch den Abstand der gestrichelten Linie ($-e\varphi$) von der strichpunktierten Linie (E_F) bestimmt. Am Schnittpunkt beider Linien ist $n = p = n_i$, links davon ist $p > n_i > n$ und rechts davon $p < n_i < n$.

5.4.2.3 Feldzone bei äußerer Spannung. Wird an den beiden Kristallenden eine äußere Spannung angelegt, so fließt durch den PN-Kristall ein Strom, der grundsätzlich das BOLTZMANN-Gleichgewicht der Stromträger verändert. Man kann aber zeigen, daß diese Abweichung vom BOLTZMANN-Gleichgewicht klein ist. Die üblicherweise durch einen Kristall fließenden Ströme sind nämlich klein gegenüber dem Diffusions- und Feldstrom, die im stromlosen Fall schon vorhanden sind und sich genau kompensieren. Deshalb lassen sich die Dichteverteilungen der Stromträger in der Feldzone so bestimmen, als ob kein Strom, also Gleichgewicht zwischen Diffusions- und Feldstrom, vorhanden wäre. Es sind also nach wie vor die Löcher in der Feldzone im BOLTZMANN-Gleichgewicht mit dem P-Gebiet (Gl. 5.5) und die Elektronen in der Feldzone mit dem N-Gebiet (Gl. 5.6).

Wegen der relativen Kleinheit des Stromes und der Trägerverarmung in der Feldzone (hoher Widerstand der Feldzone) fällt fast die gesamte äußere Spannung über der Feldzone ab. Dann bedeutet eine äußere Spannung aber nur, daß das Potential V_D um die Größe der äußeren Spannung verändert wird.

Legen wir an das P-seitige Ende ein positives äußeres Potential U, so wird in Abb. 5.23 das linke Potentialniveau um U gehoben, während das Potential der rechten äußeren Hälfte seinen alten Wert beibehält. Da von links nach rechts jetzt nur ein Potentialunterschied $V_D - U$ zu überwinden ist, sind weniger positive und negative Raumladungen notwendig als im stromlosen Fall. Die Feldzone wird, wie in Abb. 5.26 b angedeutet, schmäler; ist $U = V_D$, so ist sie sogar ganz verschwunden. Umgekehrt bedingt ein negatives Potential der P-Seite (Abb. 5.26 c) eine Feldzonenverbreiterung, weil jetzt die Raumladungen einen größeren Potentialunterschied ausgleichen müssen. Das kann aber nur durch eine größere Feldzonenbreite geschehen, weil als Raumladungen im wesentlichen nur die konstanten Akzeptoren- und Donatorendichten in Frage kommen.

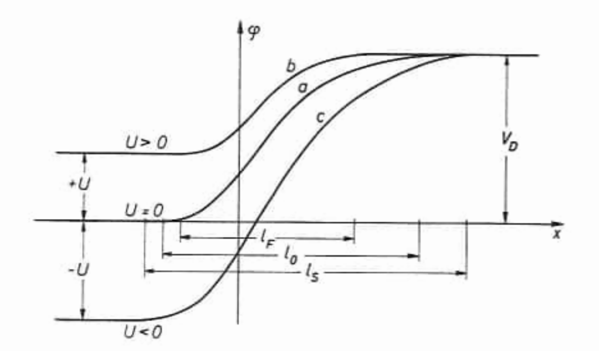


Abb. 5.26.
Potentialverlauf im PN-Übergang.
a) bei Gleichgewicht, b) bei Flußspannung, c) bei Sperrspannung.

Um sofort die Wirkungen der an- oder abgebauten Raumladungen auf die Zonenbreite zu übersehen, macht man sich bei eindimensionalen Problemen auf Grund der Poissonschen Gleichung leicht klar, daß der Beitrag einer Raumladung zur Feldstärke am PN-Sprung bei $x = 0$ gleich dem Integral der Ladungsdichte über eine Kristallhälfte ist. Der Beitrag einer solchen Raumladung zum Potential dagegen ist gleich dem Moment der Ladung, wobei jedes Ladungselement mit seinem Abstand vom PN-Sprung zu werten ist. Dieselbe Zusatzladung trägt demnach zur Feldstärke bei $x = 0$ immer denselben Beitrag bei, gleichgültig, wo sie angebracht wird, ob bei kleinen oder großen x-Werten. Dagegen ist der Beitrag einer solchen Ladung zum Potential bei $x = 0$ bei größerem Abstand größer, als wenn sie in kleinerem Abstand angebracht würde. So ist verständlich, daß die in unserem Beispiel große Ladungsdichte auf der P-Seite auf kurzen Strecken dieselbe Feldstärke bei $x = 0$ ergibt wie die kleinere Ladungsdichte auf größeren Strecken der N-Seite. Der Beitrag der gleichen Gesamtladung der N-Seite zum gesamten Potential der Doppelschicht ist wegen der größeren Entfernungen vom PN-Sprung viel größer als der Potentialbeitrag derselben Gesamtladung der P-Seite.

Die Bestimmungsgrößen der Feldzone sind ohne weitere Rechnung aus Gl. 5.19 bis 5.21 bzw. aus Gl. 5.21 a zu entnehmen. Wir brauchen dort nur V_D durch $V_D - U$ zu ersetzen, wobei $U > 0$ ist, wenn die P-Seite ein positives Potential hat, und $U < 0$ bei negativer Spannung an der P-Seite. Später wird gezeigt, daß $U > 0$ dem Durchlaßfall und $U < 0$ dem Sperrfall einer PN-Schicht entspricht. Bei wachsender Durchlaßspannung an der PN-Schicht wird also die Feldzone kleiner, wachsende Sperrspannungen verbreitern sie. Aus Gl. 5.21 erhält man für die Sperrschichtausdehnung des abrupten Übergangs bei der äußeren Spannung U

$$l = \left[\frac{2\varepsilon\varepsilon_0}{e} \frac{N_D + N_A}{N_D N_A} (V_D - U)\right]^{\frac{1}{2}} \equiv l_o \left(1 - \frac{U}{V_D}\right)^{\frac{1}{2}} \qquad \text{(Gl. 5.30)}$$

wobei l_o die Dicke der Raumladungszone ohne äußere Spannung darstellt. Bei 300 Volt Sperrspannung hat demnach die Sperrschicht des in Abb. 5.22 dargestellten PN-Übergangs eine Ausdehnung von etwa $25\,\mu$. Da V_D kleiner ist als 1 Volt, kann es bei größeren Sperrspannungen vernachlässigt werden. Die Feldzonenbreite wächst dann mit der Wurzel aus der Sperrspannung.

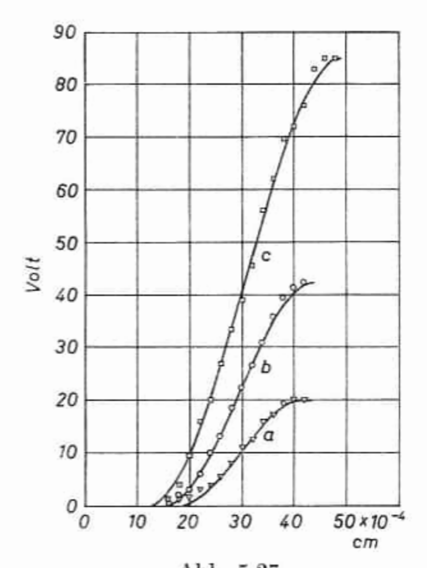


Abb. 5.27.
Gemessener Potentialverlauf in einem PN-Übergang bei verschiedenen Sperrspannungen.

a) $U = -20$ V, $l = 24\,\mu$; b) $U = 43$ V, $l = 29\,\mu$; c) $U = -85$ V, $l = 37\,\mu$ (nach PEARSON, READ und SHOCKLEY [44]).

Abb. 5.27 zeigt den mit einer feinen Wolframspitze abgetasteten Potentialverlauf in der Feldzone eines PN-Kristalls [44]. Die Sperrspannungen betragen 20, 43 und 85 Volt. Man sieht sehr schön, daß die Feldzonenbreite anwächst und der gesamte Potentialunterschied in der Feldzone liegt. Letzteres läßt sich auch mit einem einfachen Vorlesungsexperiment in der Weise zeigen, daß man mit Hilfe einer Mikrometerschraube eine Nadel über einen PN-Kristall führt und den Potentialsprung mit Hilfe eines Röhrenvoltmeters mißt.

Die Feldstärke im abrupten PN-Übergang wächst nach Gl. 5.11 von den Rändern der Feldzone linear bis zu ihrem Maximalwert $F_{\max}$ bei $x = 0$. $F_{\max}$ ist in einfacher Weise mit dem Potentialabfall und der Sperrschichtausdehnung des PN-Überganges verknüpft.

Aus Gl. 5.11, 5.19 und 5.30 ergibt sich

$$F_{\max} = \frac{2(V_D - U)}{l} = \frac{2}{l_o} \sqrt{V_D} \sqrt{V_D - U} \qquad \text{(Gl. 5.31)}$$

Für den Übergang der Abb. 5.21 erhält man mit den angegebenen Werten:

$$\text{für } U = 0 \qquad : F_{\max} = 7 \cdot 10^3 \text{ V/cm}$$

$$\text{für } U = -300 \text{ V}: F_{\max} = 2 \cdot 10^5 \text{ V/cm}$$

Der zuletzt genannte Wert kommt der Durchbruchsfeldstärke im Ge schon sehr nahe.

5.4.2.4 Kapazität der Feldzone. Die Veränderung der Sperrschichtdicke mit der angelegten Spannung führt zu einer Wechselstromkapazität. Sie wird mit einer kleinen Wechselspannung gemessen und ändert sich, wenn man den *PN*-Übergang mit verschiedenen Gleichspannungen vorspannt. Die Wirkung der der Gleichspannung überlagerten Wechselspannung ist, daß an den Enden der Feldzone kleine Ladungsdichten auf- und abgebaut werden. Die Raumladung innerhalb der Feldzone wird dadurch nicht verändert. Man hat deshalb denselben Zusammenhang zwischen Ladung und Potentialänderung wie bei einem Plattenkondensator mit dem Plattenabstand l. Die Wechselstromkapazität pro Flächeneinheit ist also (in Farad/cm²):

$$C_F = \frac{\varepsilon \varepsilon_0}{l} \qquad \text{(Gl. 5.32)}$$

Nun wächst aber l mit der Sperrspannung, z. B. beim abrupten Übergang mit der Quadratwurzel. Also nimmt die Kapazität mit zunehmender Sperrspannung entsprechend ab.

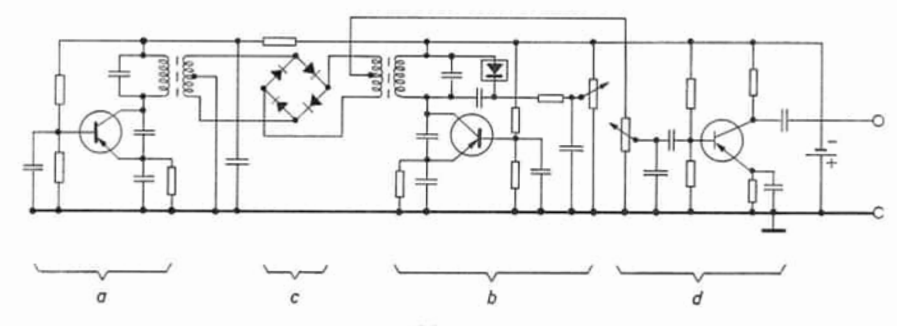


Abb. 5.28.
Schaltung zur Demonstration der variablen Kapazität einer Diode
a) Oszillator mit fester Frequenz, b) Oszillator mit variabler Frequenz; Varicap liegt parallel zum Schwingkreis. c) Ringdemodulator zur Erzeugung der Differenzfrequenz. d) Verstärker.

Diese mit der Spannung veränderliche Kapazität einer *PN*-Schicht wird, wie in Abschnitt 5.2.4 besprochen, bei der Kapazitätsdiode (Varicap) ausgenützt. Sie läßt sich im Vorlesungsversuch demonstrieren, wenn man sie als Teilkapazität zu einer Schwingkreiskapazität schaltet und diesem Kreis einen zweiten mit fester Kapazität überlagert. Die Frequenz des Überlagerungstons läßt sich allein durch die Vorspannung der *PN*-Schicht verändern. Abb. 5.28

zeigt die Schaltskizze des Überlagerungsteils, dem ein einfacher dreistufiger Verstärker mit Lautsprecher zur Demonstration vor einem größeren Zuhörerkreis nachgeschaltet wird.

Für die einfachen Fälle des abrupten und des linearen Übergangs ist die Gl. 5.30 bzw. die modifizierte Gl. 5.21 a in Gl. 5.32 einzusetzen. Gl. 5.32 gilt aber allgemein für jeden Störstellendichteverlauf. Man kann deshalb eine Kapazitätsmessung dazu benutzen, um die Veränderung der Feldzonenbreite mit der Spannung zu ermitteln und daraus den Verlauf komplizierter Störstellenprofile gewinnen.

5.4.2.5 Die Trägerdichteverteilung in der Feldzone bei äußerer Spannung. Nach den Erörterungen im Abschnitt 5.4.2.3 läßt sich dieser Fall auf den stromlosen Fall zurückführen, indem das Diffusionspotential V_D durch $V_D - U$ ersetzt wird. Beispielsweise kann dann der Potentialverlauf in der Feldzone bei Stromdurchgang, wie in Abb. 5.26 geschehen, berechnet werden. Aus demselben Grund kann auch die BOLTZMANN-Verteilung in der Feldzone als annähernd gültig angenommen werden. Es können also aus der Potentialverteilung von Abb. 5.26 gemäß Gleichung 5.5 und 5.6 die Dichten der Löcher und Elektronen in der Feldzone gewonnen werden. Das Ergebnis ist in Abb. 5.29 dargestellt.

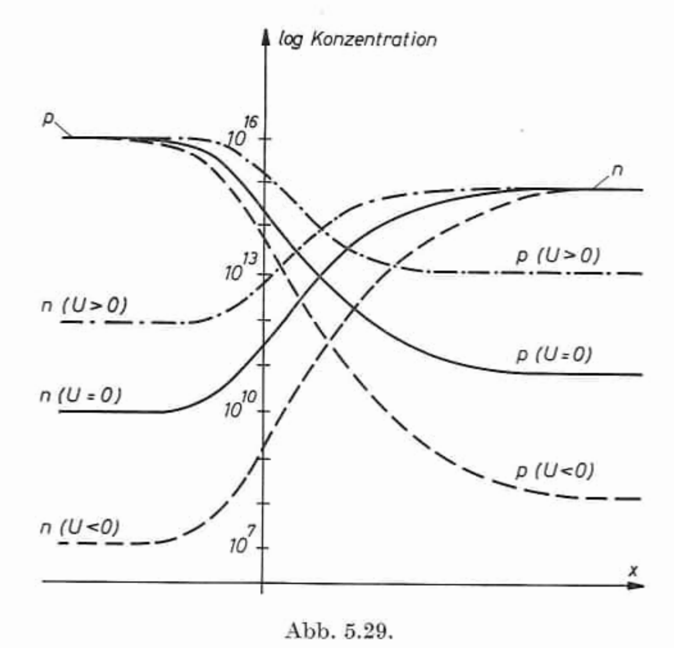


Abb. 5.29.
Löcher- und Elektronendichte in der Feldzone
(—— Gleichgewicht, –.–.–. Flußspannung, – – – Sperrspannung).

Insbesondere ist am rechten Rand x_N der Feldzone bei dem Potential U der P-Seite

$$p(x_N) = p_P \cdot e^{-(V_D - U)/U_T} \qquad \text{(Gl. 5.33)}$$

Wegen Gl. 5.2 wird dann

$$p(x_N) = p_N \cdot e^{U/U_T} \qquad \text{(Gl. 5.34)}$$

In gleicher Weise ergibt sich für die Elektronendichte am linken Rand x_P der Feldzone

$$n(x_P) = n_P \, e^{U/U_T} \qquad \text{(Gl. 5.35)}$$

Ist, verglichen mit der N-Seite, die P-Seite auf positivem Potential ($U > 0$), so sind an den Rändern der Feldzone die Minoritätendichten größer als im weit vom PN-Übergang entfernten Halbleiterinneren. Ist die P-Seite negativ ($U < 0$), so sind am gleichen Ort die Minoritätendichten kleiner als im Halbleiterinneren.

5.4.3 Diffusionszone

5.4.3.1 Allgemeines. Wir haben gezeigt, daß fast die gesamte Potentialdifferenz, die am PN-Kristall liegt, in der Feldzone abfällt. Demnach sind die an die Feldzone anschließenden Gebiete annähernd feldfrei. Ein geringes elektrisches Feld muß natürlich auch dort vorhanden sein, um die Ströme durch den Halbleiter zu treiben. Im allgemeinen ist dieses Feld jedoch klein im Vergleich zu den in der Feldzone vorhandenen Feldern. Insbesondere aber sind diese Gebiete im Gegensatz zur Feldzone raumladungsfrei.

Die bei äußerer Spannung an den Feldzonenrändern von der Gleichgewichtsdichte abweichenden Minoritätendichten suchen sich an die Gleichgewichtsdichten des Halbleiterinneren anzugleichen. Ist z. B. die P-Seite positiv (Flußrichtung), so fließen Defektelektronen außerhalb der Feldzone nach rechts und Elektronen nach links. Da außerhalb der Feldzone für den Strom der Minoritäten kein merkliches elektrisches Feld vorhanden ist, kommt als treibende Kraft im wesentlichen lediglich eine Diffusion der Minoritäten infolge des Konzentrationsunterschiedes vom Feldzonenrand nach dem Kristallinneren in Betracht. Wie ausgedehnt diese Diffusionszonen sind, wird noch erörtert. Sie beginnen an den Enden der Feldzone. In Abb. 5.31 sind die Trägerdichten von Abb. 5.29 noch einmal mit kleinerem x-Maßstab gezeichnet, um die Diffusionszonen mit auf das Bild zu bekommen.

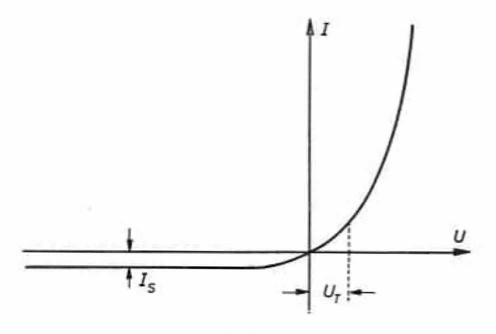


Abb. 5.30.
Ideale Strom-Spannungs-Kennlinie eines PN-Übergangs.

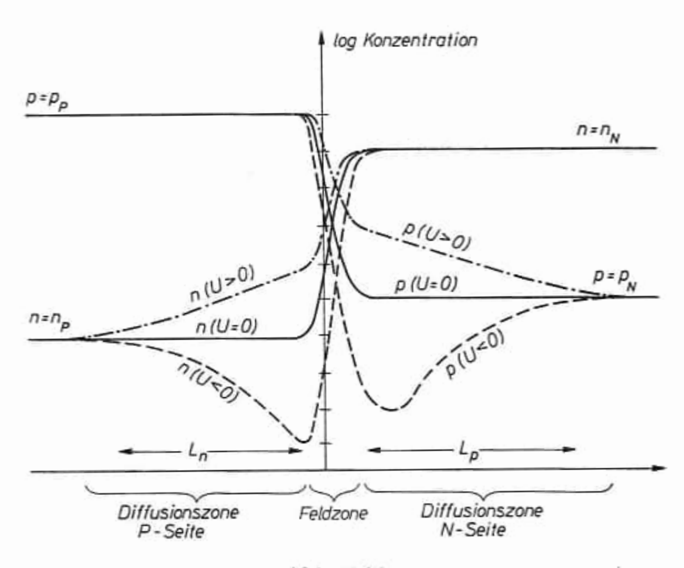


Abb. 5.31.
Löcher- und Elektronendichte in den Diffusionszonen (Abszissenmaßstab gegenüber Abb. 5.29 verkleinert).

Es wird sich zeigen, daß diese Diffusionsströme die Strom-Spannungs-Charakteristik bestimmen. Die elektrische Asymmetrie des PN-Kristalls ist dann sofort zu verstehen, weil bei positiver P-Seite ($U > 0$) die Minoritätendichten mit wachsender Spannung am Feldzonenrand immer mehr über den Gleichgewichtsdichten liegen. Man erwartet demnach mit wachsender Spannung wachsende Ströme (Abb. 5.30). Ist hingegen die P-Seite negativ ($U < 0$), so ist bereits für Spannungen von einigen 100 mV und darüber die Minoritätendichte an den Feldzonenrändern klein gegenüber der Gleichgewichtsdichte (Gl. 5.34 und 5.35); ohne merklichen Fehler kann man sie dann zu Null annehmen. (Man beachte den logarithmischen Maßstab der Konzentration in Abb. 5.31.) Die Diffusion der Minoritäten erfolgt jetzt von dem Kristallinnern nach der Feldzone. Man erhält deshalb einen konstanten Diffusionsstrom für alle Spannungen größer als 100 mV. Positive Spannungen auf der P-Seite ergeben demnach das Durchlaßverhalten, negative Spannungen das Sperrverhalten des PN-Kristalls.

5.4.3.2 Ströme in der Diffusionszone bei Stromträgerneutralisation. Die folgenden Überlegungen lassen sich am anschaulichsten für den Durchlaßfall ($U > 0$) diskutieren. Sie sind jedoch auch auf den Sperrfall ($U < 0$) anzuwenden. Für $U > 0$ ist die Löcherdichte am rechten Feldzonenrand $> p_N$, demnach findet eine Diffusion von Löchern nach rechts in die anschließenden Bereiche des N-Teils statt (Diffusionszone). Jedes Defektelektron, das in die Diffusionszone eintritt, wird nach den Erörterungen im

Abschnitt 4.2.3 sofort durch ein Elektron neutralisiert, das über den N-Teil des Kristalls zugeführt wird. Da Paare von Defektelektronen und Elektronen gemeinsam diffundieren, ist zunächst nicht einzusehen, warum der Diffusionsstrom der Löcher am Feldzonenrand einen elektrischen Strom, nämlich den Durchlaßstrom, ergibt. Die Zusatzdichte der Löcher und Elektronen nimmt mit wachsender Entfernung vom Feldzonenrand, also mit wachsendem Diffusionsweg, ab wegen der die Generation überwiegenden Rekombination. Für die Strombilanz scheiden diese diffundierenden Paare als Stromträger zwar aus, jedoch bestimmt die Geschwindigkeit der Abdiffusion der Paare vom Feldzonenrand den Strom der aus der Feldzone in die Diffusionszone eintretenden Löcher. Damit wird durch sie auch der Strom der Elektronen, der zur Neutralisation dieser Löcher aus dem N-Gebiet zufließen muß, bestimmt. Dieser Elektronenstrom setzt den positiven Strom der vom P-Teil über die Feldzone kommenden Defektelektronen nach rechts fort und ist gerade so groß wie der Diffusionsstrom der Löcher am Feldzonenrand. Man kann also so argumentieren, als ob der Durchlaßstrom von den abdiffundierenden Löchern herrührt. Falsch ist diese Vorstellung auch deshalb nicht, weil die (zur Neutralisation der austretenden Defektelektronen) von rechts an die Feldzonengrenzen geführten Elektronen und die von dort wieder nach rechts zurückdiffundierenden Elektronen sich bei der Betrachtung herausheben.

Diese Überlegung soll nun an Hand der Bilanzgleichungen untermauert werden. Der Gesamtstrom in der Diffusionszone setzt sich aus dem Strom der Elektronen und Löcher zusammen. Die Gesamtstromdichte ist

$$\boldsymbol{j} = \boldsymbol{j}_n + \boldsymbol{j}_p \qquad \text{(Gl. 5.36)}$$

Durch Subtraktion der Gleichungen 4.55 und 4.56 ergibt sich daraus für den Fall der Neutralisation, wo $\delta p = \delta n$ ist

$$\operatorname{div} \boldsymbol{j} = 0 \qquad \text{(Gl. 5.37)}$$

d. h. der elektrische Gesamtstrom ist quellenfrei. Dieses Ergebnis wird verständlich, wenn man bedenkt, daß Generation und Rekombination paarweise erfolgten und deshalb nicht als Quellen oder Senken des elektrischen Stromes in Frage kommen. In dem von uns betrachteten eindimensionalen Fall folgt daraus, daß die Gesamtstromdichte

$$j = \text{const} \qquad \text{(Gl. 5.38)}$$

ist.

Es soll nun untersucht werden, zu welchem Anteil die Elektronen und Löcher an den einzelnen Stellen der Diffusionszone am Gesamtstrom beteiligt sind, und in welchem Verhältnis bei den einzelnen Stromträgern Feld- und Diffusionsstrom stehen. Wir haben schon erwähnt, daß auch in der Diffusionszone eine, wenn auch geringe, örtlich eventuell unterschiedliche elektrische

Feldstärke vorhanden ist. Diese, zusammen mit den örtlich vorhandenen Konzentrationsgefällen von Elektronen und Defektelektronen, bestimmen Elektronen- und Löcherstrom, für die unabhängig voneinander die Stromgleichungen 4.49 und 4.50 gelten. Durch Addition der beiden Stromgleichungen folgt für das Feld $\boldsymbol{F}$ unter Berücksichtigung von Gl. 5.36:

$$\boldsymbol{F} = -\operatorname{grad} \varphi = \frac{\boldsymbol{j}}{\sigma} - \frac{e(D_n - D_p)}{\sigma} \operatorname{grad} \delta p \qquad \text{(Gl. 5.39)}$$

Dabei ist die Leitfähigkeit $\sigma = e(n\mu_n + p\mu_p)$ eingeführt und, wegen $\delta p = \delta n$, $\operatorname{grad} n = \operatorname{grad} p = \operatorname{grad} \delta n$ gesetzt. $\boldsymbol{F}$ besteht aus zwei Teilen: einmal ist ein Feld nötig, um den Strom j durch die Diffusionszone zu treiben; außerdem verursachen die verschiedenen Diffusionskonstanten ein elektrisches Zusatzfeld, das nötig ist, um bei beiden Partnern *dieselbe* Diffusionsgeschwindigkeit zu erzwingen. Die schneller diffundierenden Elektronen ziehen die langsamer diffundierenden Löcher hinter sich her und werden selbst dadurch gebremst. Einsetzen von Gleichung 5.39 in 4.49 und 4.50 ergibt

$$\boldsymbol{j}_n = \frac{\sigma_n}{\sigma} \boldsymbol{j} + e\, D^* \operatorname{grad} \delta p \qquad \text{(Gl. 5.40)}$$

$$\boldsymbol{j}_p = \frac{\sigma_p}{\sigma} \boldsymbol{j} - e\, D^* \operatorname{grad} \delta p \qquad \text{(Gl. 5.41)}$$

mit

$$D^* = \frac{\sigma_n D_p + \sigma_p D_n}{\sigma} \qquad \text{(Gl. 5.42)}$$

$$\sigma_n = e\, n\, \mu_n \qquad \text{(Gl. 5.43)}$$

und

$$\sigma_p = e\, p\, \mu_p \qquad \text{(Gl. 5.44)}$$

D^* ist die Diffusionskonstante, auf die sich Löcher und Elektronen schließlich einigen.

Für einen N-Leiter ($n \gg p$) ist $D^* \approx D_p$, für einen P-Leiter $D^* \approx D_n$. Hier ist also jeweils die Diffusionskonstante der Minoritäten bestimmend. In unserem Fall der Diffusion im N-Gebiet ist $\sigma_p \ll \sigma_n \approx \sigma$. Die Gleichungen 5.40 und 5.41 vereinfachen sich dann zu

$$\boldsymbol{j}_n = \boldsymbol{j} + e\, D_p \operatorname{grad} \delta p \qquad \text{(Gl. 5.45)}$$

$$\boldsymbol{j}_p = \quad -e\, D_p \operatorname{grad} \delta p \qquad \text{(Gl. 5.46)}$$

$\boldsymbol{j}$ ist nach Gl. 5.38 eine Konstante. Man kann für die Diffusion im N-Teil des Schichtkristalls das Schema der Abb. 5.32 angeben, wobei wir einen positiven Strom nach rechts annehmen, was dem Durchlaßfall entspricht. Am Feldzonenrand bei $x = x_N$ strömen nur Löcher aus der Feldzone. Es kommen keine Elektronen aus der Feldzone, d. h. nach Gl. 5.45, daß dort $j_n = 0$, also der

Diffusionsstrom der Elektronen negativ gleich dem Feldstrom j der Elektronen ist. Daraus folgt weiter, daß der Löcherdiffusionsstrom bei $x = x_N$ gleich dem Gesamtstrom ist. Mit wachsendem x nimmt grad δp ab (Abb. 5.31, beachte den logarithmischen Maßstab!). Der Löcherstrom wird also kleiner, der Elektronenstrom größer, bis für große x-Werte grad δp sehr klein, der gesamte Strom j also ein Elektronenstrom geworden ist. Wir haben uns oben auf die Diffusion von Löchern im N-Teil des Kristalls beschränkt. Dazu kommt jetzt noch die Diffusion von Elektronen von der Feldzone in den P-Teil des Kristalls hinein, was in entsprechender Weise behandelt werden könnte. Beide Ströme sind dann zu addieren, um den gesamten Strom durch den PN-Übergang zu erhalten. Außerdem läßt sich das Gesagte statt auf den Durchlaßfall auch auf den Sperrfall anwenden, was auch unterlassen sei. Wir haben aber jetzt genügend klar dargelegt, daß die Summe der Diffusionsströme der Minoritäten an den Rändern der Feldzone gleich dem Gesamtstrom durch die PN-Schicht ist.

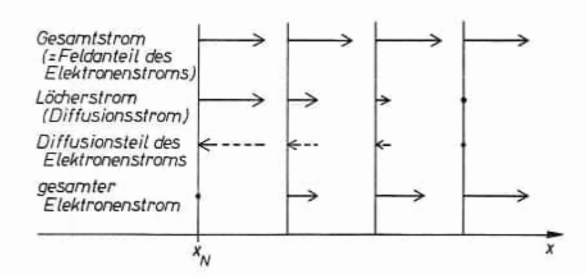


Abb. 5.32.
Aufteilung der Ströme im N-Teil eines PN-Überganges bei Flußspannung. (Die dargestellten Ströme sind elektrische Ströme, nicht Teilchenströme.)

5.4.3.3 Berechnung des Sättigungsstromes. Als einfachsten Teil der Strom-Spannungs-Kennlinie berechnen wir zunächst den Sättigungsstrom in Sperrichtung. Auf der N-Seite diffundieren dann Defektelektronen nach links zum Feldzonenrand, wo die Defektelektronendichte nahezu Null ist (Abb. 5.31 und 5.33). Ebenso diffundieren die Elektronen im P-Teil des Kristalls nach rechts. Für die Diffusion der Löcher im N-Gebiet verwenden wir die Bilanzgleichung in der Form von Gl. 4.75, wobei wir g_L und $\partial p/\partial t$ gleich Null setzen (keine Beleuchtung und stationärer Fall, d. h. Gleichstrom). Für j_p setzen wir den Ausdruck von Gl. 5.46 ein. Wir erhalten dann für unseren eindimensionalen Fall

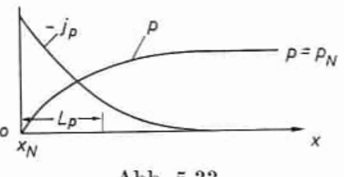


Abb. 5.33.
Verlauf der Löcherdichte p und des Löcherstromes in der Diffusionszone des N-Gebiets bei Sperrspannung. (Im Gegensatz zu Abb. 5.31 ist p in linearem Maßstab aufgetragen.).

$$\frac{\delta p}{\tau_p} - D_p \frac{\partial^2 \delta p}{\partial x^2} = 0 \qquad \text{(Gl. 5.47)}$$

Die Randbedingungen sind: $p(x_N) = 0: \quad \delta p(x_N) = -p_N$ (Gl. 5.48)

$p(\infty) = p_N: \quad \delta p(\infty) = 0$ (Gl. 5.49)

Die allgemeine Lösung von Gl. 5.47 ist bei konstantem τ_p

$$\delta p = C_1\, e^{x/L_p} + C_2\, e^{-x/L_p} \tag{Gl. 5.50}$$

wobei

$$L_p \equiv \sqrt{D_p \tau_p} \tag{Gl. 5.51}$$

ist.

Wegen der Randbedingung Gl. 5.49 ist $C_1 = 0$; und wegen Gl. 5.48 ist

$$C_2 = -\, p_N\, e^{x_N/L_p} \tag{Gl. 5.52}$$

damit wird

$$p = p_N + \delta p = p_N\left(1 - e^{-(x-x_N)/kT}\right) \tag{Gl. 5.53}$$

Die Löcherdichte fällt also nach einer Exponentialfunktion vom Halbleiterinneren nach dem Rand der Feldzone ab (Abb. 5.33). Die charakteristische Länge L_p dieses Abfalls heißt Diffusionslänge oder Diffusionsweg der Löcher. Es ist dies die Länge, die von einem diffundierenden Loch im Mittel zurückgelegt werden kann, bevor es rekombiniert. Die Ausdehnung der Diffusionszone ins N- und P-Gebiet hinein ist also von der Größenordnung L_p bzw. L_n (siehe Abb. 5.31).

Aus Gl. 5.53 gewinnt man mit Gl. 5.46 die Diffusionsstromdichte der Löcher

$$j_p = -\,\frac{e\, D_p\, p_N}{L_p}\, e^{-(x-x_N)/kT}$$

Dieser Verlauf ist ebenfalls in Abb. 5.33 eingezeichnet. Der Sättigungsstrom ist nach dem in Abschnitt 5.4.3.2 Gesagten

$$j_{ps} = -\, j_p\big|_{x=x_N} = e\, D_p\, \frac{p_N}{L_p} \tag{Gl. 5.54}$$

Dieses Ergebnis kann folgendermaßen anschaulich gemacht werden: In der N-Diffusionszone ist im Sperrfall überall die Löcherdichte unter dem Gleichgewichtswert p_N. Deshalb überwiegt hier die Generation über die Rekombination. Die Zahl der pro cm³ und sec generierten Löcher ist (nach Gl. 4.69 und 4.72) p_N/τ_p. Diese Löcher strömen zur Löchersenke bei $x = x_N$. Deshalb nimmt der Löcherstrom nach links zu. Da aber ein generiertes Loch durch Diffusion im Mittel keinen größeren Weg als L_p zurücklegen kann, tragen nur die Löcher aus einer Schicht von der Dicke L_p zum Diffusionsstrom bei. Der am Feldzonenrand fließende Strom pro cm² ist deshalb

$$j_{ps} = e\, \frac{p_N}{\tau_p}\, L_p \tag{Gl. 5.55}$$

was wegen Gl. 5.51 der Gl. 5.54 entspricht.

Das Ergebnis Gl. 5.54 für den Sättigungsstrom der Löcher kann noch auf eine zweite Weise veranschaulicht werden, wenn man bedenkt, daß dieser Strom

ein Diffusionsstrom, also proportional zu e, D_p und dem Dichtegradienten ist. Der Gradient der Löcherdichte kann aus Abb. 5.33 zu p_N/L_p abgeschätzt werden, was wieder auf Gl. 5.54 führt.

Durch Betrachtung der Vorgänge auf der P-Seite erhält man für den Diffusionsstrom im Falle großer Sperrspannungen einen der Gl. 5.54 entsprechenden Ausdruck für die Elektronen

$$j_{ns} = e\, D_n \frac{n_P}{L_n} \qquad \text{(Gl. 5.56)}$$

Die gesamte Sättigungssperrstromdichte ist dann

$$j_s = \frac{e\, D_p\, p_N}{L_p} + \frac{e D_n\, n_p}{L_n} \qquad \text{(Gl. 5.57)}$$

Technische Gleichrichter bestehen meist aus stark asymmetrischen PN-Übergängen, so daß eines der beiden Glieder von Gl. 5.57 in der Regel vernachlässigt werden kann.

Die zunächst überraschende Tatsache, daß die Minoritäten den Sperrstrom — überhaupt das Gleichrichterverhalten — einer PN-Schicht bestimmen, läßt sich einfach demonstrieren, wenn die Sperrströme von PN-Übergängen in Material von verschiedenen spezifischen Widerständen verglichen werden, z. B. von einem mit In legierten PN-Gleichrichter mit niederohmigem und hochohmigem N-Ge. Solche Kennlinien sind in Abb. 5.34 schematisch aufgezeichnet. Der Gleichrichter mit der größeren Elektronendichte hat den kleineren Sperrsättigungsstrom. Dies entspricht der Beziehung

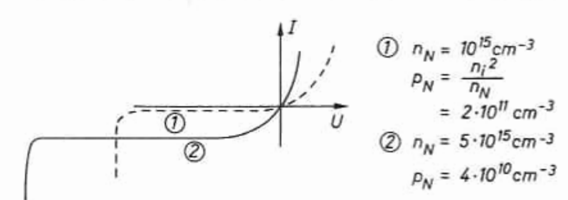


Abb. 5.34.
PN-Kennliniebei hoher (1) und niedriger (2) Dotierung.

$$j_{ps} \sim p_N = \frac{n_i^2}{n_N} \approx \frac{n_i^2}{N_D} \qquad \text{(Gl. 5.58)}$$

Auch die Temperaturabhängigkeit des Sättigungsstromes läßt sich zur experimentellen Bestätigung der Gl. 5.54 bis 5.58 heranziehen. Gemäß Gl. 5.1 muß n_i^2 und damit auch der Sättigungsstrom mit wachsender Temperatur ansteigen. Dies ist in Abb. 5.35 für einen derartigen legierten Germaniumgleichrichter gezeigt.

Schließlich wird Gl. 5.58 noch durch den Vergleich der Sperrströme von Germaniumgleichrichtern mit Siliziumgleichrichtern bestätigt, da die Eigenleitungsdichte wegen des größeren Bandabstandes in Si nach Tabelle 1 viel

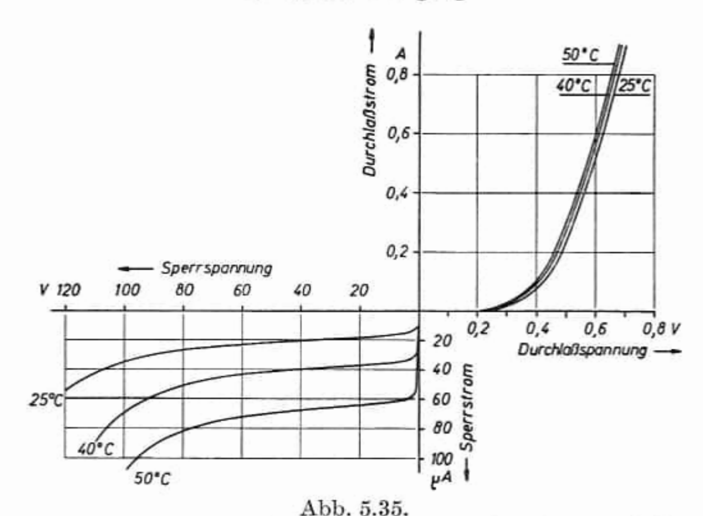


Abb. 5.35.
Kennlinien eines Germanium-Flächengleichrichters bei verschiedenen Temperaturen (nach RAITHEL und SPITZER [47]).

kleiner ist als in Ge. Der zugehörige Unterschied im Sperrstrom ist in Abb. 5.2 in sehr eindrucksvoller Weise sichtbar. Im übrigen folgt der PN-Übergang in Silizium, wie schon in Abschnitt 5.1 angedeutet, nicht vollständig der hier ausgeführten einfachen Theorie. Dies ist auf eine erhöhte Rekombinationsrate in der Feldzone auf Grund von besonderen Rekombinationszentren zurückzuführen. (Näheres siehe KLEINKNECHT und SEILER [9] und SAH, NOYCE und SHOCKLEY [36]).

5.4.3.4 Gleichstromkennlinie des PN-Übergangs. Im vorhergehenden Abschnitt wurde der Sättigungsstrom in Sperrichtung aus der Bilanzgleichung Gl. 4.75 (für Dunkelheit $g_L = 0$ und Gleichstrom $\partial p/\partial t = 0$, d. h. aus der Gl. 5.47) dadurch gewonnen, daß die Minoritätendichte am Feldzonenrand gleich Null gesetzt wurde (Gl. 5.48). Wir wollen nun den Strom durch den PN-Übergang für beliebige Spannungen U berechnen und haben Gl. 5.48 deshalb durch Gl. 5.34 zu ersetzen:

$$\delta p(x_N) = p_N \left(e^{U/U_T} - 1 \right) \qquad \text{(Gl. 5.59)}$$

Die andere Randbedingung Gl. 5.49 bleibt unverändert. Die allgemeine Lösung der Differentialgleichung 5.47 ist wieder durch Gl. 5.50 gegeben. Die Anwendung der neuen Randbedingung Gl. 5.59 ergibt nun

$$p = p_N + \delta p = p_N \left[1 + \left(e^{U/U_T} - 1 \right) e^{-(x - x_N)/L_p} \right] \qquad \text{(Gl. 5.60)}$$

Die Löcherstromdichte der N-Seite ist gleich der Löcherdiffusionsstromdichte

bei $x = x_N$, also

$$j_p(x_N) = - e D_p \operatorname{grad} p \,|_{x_N} = j_{ps} (e^{U/U_T} - 1) \qquad \text{(Gl. 5.61)}$$

Beim letzten Glied dieser Gleichung wurde der Ausdruck für den Sättigungsstrom von Gl. 5.54 verwendet. Für große Sperrspannungen geht Gl. 5.61 in Gl. 5.54 über.

Berücksichtigt man noch den Elektronenstrom der *P*-Seite, so erhält man für den gesamten Strom einer *PN*-Schicht bei der Gleichspannung U

$$j = j_s\left(e^{U/U_T} - 1\right) \qquad \text{(Gl. 5.62)}$$

Dabei ist
$$j_s = \frac{e\,D_p\,p_N}{L_p} + \frac{e\,D_n\,n_P}{L_n} = e\,n_i^2\left(\frac{D_p}{L_p\,n_N} + \frac{D_n}{L_n\,p_P}\right) \qquad \text{(Gl. 5.63)}$$

was Gl. 5.57 entspricht. Die Stromspannungsbeziehung Gl. 5.62 ist sehr einfach. Schon C. WAGNER [48] hatte sie für den Kupferoxydulgleichrichter abgeleitet, weshalb man auch öfters von der WAGNERschen Beziehung spricht. Auffallend ist, daß die meisten Theorien der Sperrschichten (mehr oder weniger falsche!) dieselbe Spannungsfunktion ergeben, während der Sättigungsstrom oft sehr verschieden herauskommt. Es ist deshalb auch die wesentlichste Leistung der hier geschilderten SHOCKLEYschen Theorie, die Größe j_s richtig zu liefern. U_T beträgt bei Zimmertemperatur nur etwa 25 mV. Man erhält deshalb in Durchlaßrichtung sehr bald (d. h. für $U > U_T = 25$ mV) einen exponentiellen Kennlinienverlauf, während in Sperrichtung sich für Sperrspannungen $-U \gg U_T$ ein konstanter Sperrstrom ergibt (vgl. Abb. 5.30).

Im vorangegangenen Abschnitt haben wir gemessene Kennlinien von legierten Gleichrichtern aus Germanium und Silizium zur qualitativen Bestätigung der Theorie herangezogen. Zu einem quantitativen Vergleich zwischen Theorie und Experiment sind gezogene *PN*-Übergänge in Stäbchenform (siehe Abb. 5.14a) besser geeignet, weil an ihnen die Majoritätenkonzentrationen p_P und n_N und die Lebensdauern τ_p und τ_n der beiden Seiten unabhängig gemessen werden können.

Abb. 5.36 zeigt das Ergebnis einer derartigen Messung von GOUCHER, PEARSON, SPARKS, TEAL und SHOCKLEY [34]. Wie man sieht, ist die Übereinstimmung über mehrere Größenordnungen hinweg sehr gut. Hierbei wurde der theoretische Wert von $U_T = 25{,}6$ mV verwendet und ein Wert für den Sättigungsstrom j_s durch Anpassung bestimmt. Der Sättigungsstrom wurde außerdem nach Gl. 5.63 berechnet. Dazu wurde n_N und p_P durch Leitfähigkeitsmessung an beiden Seiten nach Abb. 2.8 und L_p und L_n nach der Methode von Abschnitt 5.5.3 gemessen. Beide Werte stimmten innerhalb 30 % überein.

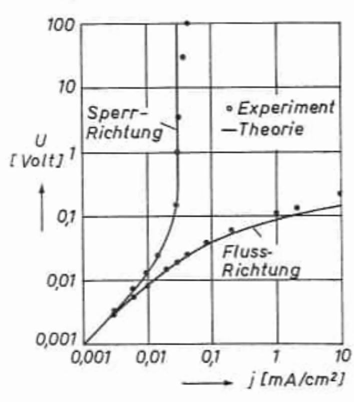


Abb. 5.36.

Kennlinie eines gezogenen Germanium-PN-Übergangs (nach GOUCHER, PEARSON, SPARKS, TEAL und SHOCKLEY [34]).

Da bei der obigen Kennlinienmessung Potentialsonden für die Spannung U direkt am PN-Übergang verwendet wurden, war kein Serienwiderstand zu berücksichtigen. Dies ist in einem größeren Temperaturbereich experimentell nicht mehr so leicht durchzuführen. Bei dem Meßergebnis von GEIST und SEILER [35] in Abb. 5.37 ist deshalb ein Serienwiderstand R_b zu berücksichtigen, der aus dem Kennlinienteil für hohe Flußspannungen gewonnen werden kann. Dann gilt für den Strom I als Funktion der äußeren Spannung U_A

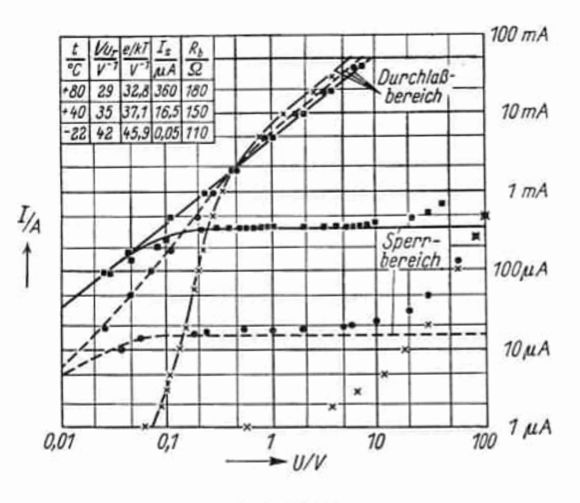


Abb. 5.37.
Kennlinie eines gezogenen PN-Übergangs bei verschiedenen Temperaturen (nach GEIST und SEILER [35]).

$$I = I_s \left(e^{(U_A - I R_b)/U_T} - 1 \right) \qquad \text{(Gl. 5.64)}$$

Die Größen $1/U_T$ und I_s sind durch Anpassung der Meßwerte an Gl. 5.64 bestimmt worden und sind in der Tabelle von Abb. 5.37 aufgeführt. Die Übereinstimmung der $1/U_T$-Werte mit dem theoretischen e/kT ist als gut zu bezeichnen. Auch bezüglich des Temperaturgangs der I_s läßt sich Übereinstimmung mit der Theorie zeigen. Oberhalb 10 Volt Sperrspannung erfolgt ein steiler Abbruch des Stromes nach oben, der ZENER-Durchbruch, der in Abschnitt 5.7.2 kurz behandelt werden wird.

5.5 Lichtelektrisches Verhalten des PN-Kristalls

5.5.1 Gleichmäßige Beleuchtung. Da das elektrische Verhalten des PN-Schichtkristalls von den Minoritätendichten bestimmt wird, ist anzunehmen, daß die Gleichrichterkennlinie durch Licht stark beeinflußbar ist. Um das Grundsätzliche zu zeigen, wollen wir uns auf den in Abb. 5.38 gezeichneten Fall einer dünnen Platte beschränken, die senkrecht zur Plattenfläche einen PN-Übergang enthält und gleichmäßig beleuchtet ist. Entsprechend den Überlegungen in Abschnitt 4.2.5 kann dann so gerechnet werden, als ob dadurch gleichmäßig im ganzen Volumen $\overline{g_L}$ Paare pro cm³ erzeugt würden.

Die in der Feldzone erzeugten Paare werden von dem dort herrschenden Feld sofort getrennt, die Elektronen werden nach rechts, die Löcher nach links gezogen. Dies ergibt einen zusätzlichen Strom in Sperrichtung. Er ist gleich der Zahl der pro cm³ und sec erzeugten Paare $\overline{g_L}$, mal dem Volumen der Feldzone $A \cdot l$, mal der Elementarladung e. Man hat also als Stromdichte $j = I/A = e\,\overline{g_L}\,l$. Wegen der Kleinheit von l ist dieser Stromanteil allerdings sehr klein. Wesentlich größer ist der Stromanteil der in der Diffusionszone erzeugten Träger. Die in der Diffusionszone, d. h. im Abstand L_p bzw. L_n vom Feldzonenrand, erzeugten Paare können nämlich zur Feldzone hindiffundieren, werden dort getrennt und liefern eine in Sperrichtung zeigende Photostromdichte von der Größe $e\,\overline{g_L}\,(L_p + L_n)$, die groß gegenüber $e\,\overline{g_L}\,l$ ist. Wir brauchen also nur diesen aus der Diffusionszone stammenden Strom zu betrachten.

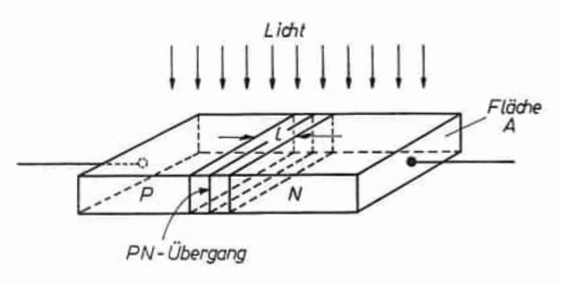


Abb. 5.38.
Photoeffekt an einem PN-Übergang bei gleichmäßiger Beleuchtung.

Um diesen Photostrom exakt abzuleiten, gehen wir von der vollständigen Bilanzgleichung 4.75 der Löcher auf der N-Seite aus, wobei wir wieder für unseren stationären Fall $\frac{\partial p}{\partial t} = 0$ setzen und Gl. 5.46 benützen:

$$0 = \overline{g_L} - \frac{\delta p}{\tau_p} + D_p \frac{\partial^2 \delta p}{\partial x^2} \qquad \text{(Gl. 5.65)}$$

Als Randbedingung haben wir für $x \to \infty$

$$\left.\frac{\partial \delta p}{\partial x}\right|_{x=\infty} = 0 \qquad \text{(Gl. 5.66)}$$

Und am Feldzonenrand ist die Löcherdichte wie im Fall ohne Beleuchtung durch das BOLTZMANN-Gleichgewicht mit der P-Seite gegeben (Gl. 5.34), da dieses aus den gleichen Gründen, wie in Abschnitt 5.4.2.3 und 5.4.2.5 aufgeführt, durch die relativ kleinen Photoströme nicht verändert wird. Es ist also die zweite Randbedingung

$$\delta p(x_N) = p_N\,(e^{U/U_T} - 1) \qquad \text{(Gl. 5.67)}$$

Somit ergibt sich als Lösung:

$$\delta p = \left[p_N \left(e^{U/U_T} - 1 \right) - \overline{g_L}\,\tau_p \right] e^{-(x-x_N)/L_p} + \overline{g_L}\,\tau_p \qquad \text{(Gl. 5.68)}$$

und damit

$$j_p = -e D_p \frac{\partial \delta p}{\partial x}\bigg|_{x_N} = \frac{e D_p p_N}{L_p}(e^{U/U_T} - 1) - \frac{e D_p \overline{g_L} \tau_p}{L_p}$$

$$= j_{ps}(e^{U/U_T} - 1) - e \overline{g_L} L_p \qquad \text{(Gl. 5.69)}$$

Das entsprechende Gleichungssystem, auf die P-Seite angewandt, liefert den Elektronenstrom, und man erhält als Gesamtstrom

$$j = j_s(e^{U/U_T} - 1) - e \overline{g_L}(L_p + L_n) = j_{\text{dunkel}} + j_L \qquad \text{(Gl. 5.70)}$$

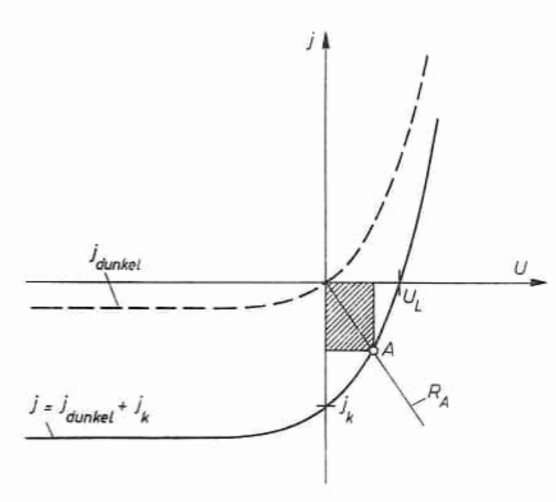


Abb. 5.39.
Kennlinie eines beleuchteten PN-Übergangs.

Hier ist j_{dunkel} die Stromdichte des unbeleuchteten PN-Übergangs nach Gl. 5.62, 5.63. Dazu addiert sich nun der Photostrom j_L, der das Vorzeichen eines Sperrstromes hat und unabhängig von der Spannung am PN-Übergang ist.

Dieses Ergebnis ist in Abb. 5.39 aufgetragen; j_{dunkel} ist die gestrichelte, der Gesamtstrom j die ausgezogene Kurve. Man sieht, daß bei Vorspannung $U = 0$, bei dem j_{dunkel} verschwindet, ein Kurzschlußstrom übrigbleibt, der gleich dem Photostrom j_L ist.

$$j_k = -e \overline{g_L}(L_p + L_n) \qquad \text{(Gl. 5.71)}$$

Der Punkt $j = 0$ ergibt die Leerlaufspannung U_L, die an den offenen Klemmen unserer beleuchteten Photodiode auftritt. Sie kann aus Gl. 5.70 mit $j = 0$ gewonnen werden. Man erhält:

$$U_L = U_T \ln\left(1 + \frac{|j_L|}{j_s}\right) \qquad \text{(Gl. 5.72)}$$

Dies ergibt für einen symmetrischen PN-Übergang ($p_P = n_N$; $n_P = p_N$) mit gleichen Lebensdauern für beide Seiten ($\tau_p = \tau_n$) einen besonders einfachen

Ausdruck, wie sich mit Hilfe der Gleichungen 5.70, 5.63 und 5.51 zeigen läßt:

$$U_L = U_T \ln \frac{p_N + \overline{g_L} \tau_p}{p_N} \qquad \text{(Gl. 5.73)}$$

Diese Formel ist nun anschaulich leicht zu verstehen. Bei offenen Klemmen werden in der Feldzone so lange Trägerpaare getrennt, bis sich ein so großes Gegenfeld aufgebaut hat, daß sich Feld- und Diffusionsstrom genau kompensieren. Dann ist in der „Diffusionszone" kein grad p mehr vorhanden. Wir haben wie im Photoleiter Abschnitt 4.2.5 $p = p_N + \overline{g_L} \tau_p$. Da aber nach wie vor das Boltzmann-Gleichgewicht Gl. 5.34 gilt, folgt daraus unmittelbar Gl. 5.73.

Die Leerlaufspannung kann bei steigender Lichtintensität aber nicht unbegrenzt weiterwachsen. Es können nämlich nur so lange Trägerpaare in der Sperrschicht getrennt werden, als dort ein Feld vorhanden ist. Dieses wird jedoch mit zunehmendem U_L verkleinert (siehe Abb. 5.26 und Gl. 5.31) und verschwindet für $U_L = V_D$. Es ist somit

$$U_L \leq V_D = U_T \ln \frac{p_P n_N}{n_i^2} \qquad \text{(Gl. 5.74)}$$

und für nichtentartetes Material ($p_P < N_V$, $n_N < N_C$) wegen Gl. 4.34

$$U_L \leq V_D < \frac{E_G}{e} \qquad \text{(Gl. 5.75)}$$

Die Tatsache, daß ein Teil der Kennlinie von Abb. 5.39 im 4. Quadranten liegt, bedeutet offenbar, daß dort trotz einer äußeren Spannung in Flußrichtung ein Sperrstrom fließt. Wir haben also eine Spannungs- bzw. Stromquelle, d. h. einen aktiven Zweipol. Der Arbeitspunkt A wird durch die Lastlinie des äußeren Widerstandes R_A bestimmt. Auch die vom Photoelement an die Last R_A abgegebene Leistung (schraffiertes Rechteck in Abb. 5.39) ist eine Funktion von R_A (näheres siehe z. B. [49]).

5.5.2 Beispiel einer praktischen Photozelle: die Solarbatterie. Die Silizium-Solarbatterie ist eines der wichtigsten Beispiele einer reinen PN-Photozelle. Sie findet unter anderem als Spannungsquelle in Erdsatelliten und Raumschiffen Verwendung. Die aus dem täglichen Leben bekanntere Selenphotozelle ist in ihrer chemischen Struktur und ihrer physikalischen Wirkungsweise weniger geklärt und sei deshalb hier nur am Rande erwähnt.

Abb. 5.40 gibt die Si-Photozelle schematisch wieder. In eine P-Typ-Siliziumplatte ist oberflächlich Phosphor eindiffundiert, so daß eine N-Schicht und ein

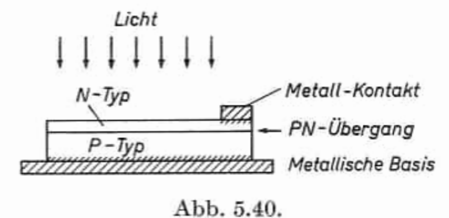


Abb. 5.40.
Schematischer Aufbau einer Solarzelle.

großflächiger PN-Übergang entsteht. Grundplatte und Schicht sind mit metallischen, ohmschen Kontakten versehen.

Die Wirkungsweise ist dieselbe wie in Abschnitt 5.5.1, wenn auch die quantitativen Verhältnisse anders sind als in dem obigen idealisierten Fall. Das Licht wird im wesentlichen in der Oberfläche der N-Schicht absorbiert, und die dort erzeugten Trägerpaare diffundieren durch die Schicht zum PN-Übergang. Um einen günstigen Wirkungsgrad zu erzielen, muß dafür gesorgt werden, daß möglichst wenig Träger rekombinieren bevor sie die Feldzone erreichen. Dazu ist eine geringe Schichtdicke und eine große Lebensdauer τ_p in der Schicht notwendig. Gleichzeitig soll aber der Querwiderstand in der Schicht klein sein, um die Verlustleistung klein zu halten. Da sich diese beiden Forderungen widersprechen, muß hinsichtlich der Schichtdicke und -dotierung ein Kompromiß gefunden werden. Die verfügbare Spannung einer solchen Zelle ist durch Gl. 5.75 auf die Breite der Bandlücke begrenzt. Aus diesem Grunde ist hier Silizium ($E_G = 1{,}2$ eV) dem Germanium ($E_G = 0{,}78$ eV) überlegen. Auch sind Versuche gemacht worden, das Silizium durch Halbleiter noch höherer Bandabstände wie z. B. Galliumarsenid ($E_G = 1{,}35$ eV) zu ersetzen [50].

5.5.3 Messung der Diffusionslänge. Der Photoeffekt eines PN-Übergangs kann dazu verwendet werden, die Diffusionslängen L_p und L_n der Minoritätsträger zu messen und damit, falls die Beweglichkeit bekannt ist, mit Hilfe von Gl. 5.51 und Gl. 4.53, 4.54 die Lebensdauer τ_p und τ_n zu bestimmen.

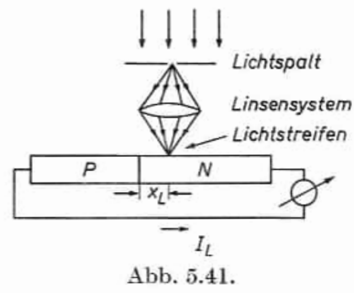


Abb. 5.41.
Bestimmung der Diffusionslänge mit Hilfe des Photoeffektes.

Abb. 5.41 gibt das Prinzip der Anordnung wieder, wie es zuerst von Goucher [51] vorgeschlagen wurde. Ein Lichtstreifen wird so auf ein PN-Stäbchen projiziert, daß er parallel zum PN-Übergang und im Abstand x_L von diesem liegt. Der gemessene Photostrom ist dann proportional zu der Zahl der vom Licht erzeugten Minoritäten, die an den PN-Übergang gelangen, ohne zu rekombinieren. Er ist wie eine Auswertung der Bilanzgleichung, Gl. 5.65 mit den Randbedingungen $\delta p(0) = 0$ und $\delta p(x_L) = \text{const.}$ ergibt:

$$j_L \sim -1/\sinh\left(\frac{x_L}{L_p}\right) \approx -e^{-x_L/L_p} \qquad \text{(Gl. 5.76)}$$

Die Näherung gilt für $x_L > L_p$. Die Messung des Photostromes bei verschiedenen Abständen x_L gibt somit die Möglichkeit einer Bestimmung von L_p. Wird der Lichtstreifen auf die P-Seite projiziert, so kann L_n bestimmt werden.

5.6 Impuls- und Wechselstromverhalten

Wir haben bisher den PN-Kristall nur unter dem Einfluß von Gleichstrom betrachtet. Dies hat sich mathematisch darin ausgedrückt, daß wir in den Bilanzgleichungen 4.75 und 4.76 das Glied $\frac{\partial}{\partial t}$ gleich Null gesetzt haben. Das Impuls- und Wechselstromverhalten der PN-Übergänge ist aber zur Vertiefung des Verständnisses besonders geeignet, weil dabei der bipolare Leitungsmechanismus stark hervortritt. Gerade dieses Verhalten ist auch für die meisten praktischen Anwendungen der PN-Schichten von entscheidender Bedeutung.

Daß hier neue Erscheinungen auftreten, zeigt schon der folgende, sehr einfache Versuch, dessen Anordnung in Abb. 5.42 dargestellt ist. An eine Germanium-PN-Diode werden sinusförmige Wechselspannungen verschiedener

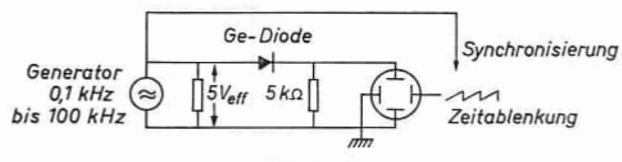


Abb. 5.42.
Schaltung zur Demonstration des Wechselstromverhaltens einer PN-Diode (nach SEILER und WUCHERER [52]).

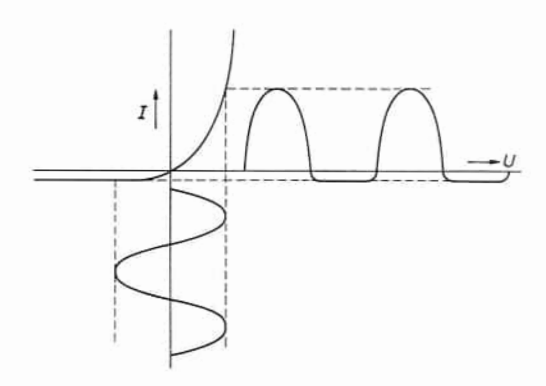


Abb. 5.43.
Gleichrichtung bei niedriger Frequenz.

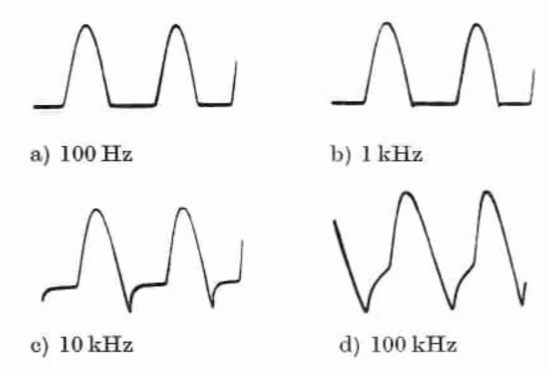


Abb. 5.44.
Wechselstromverhalten einer Ge-Flächendiode bei verschiedenen Frequenzen (nach SEILER und WUCHERER [52]).

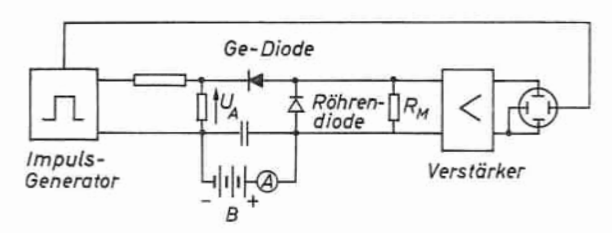


Abb. 5.45.
Anordnung zur Messung der Sperrträgheit (nach SEILER und WUCHERER [52]).

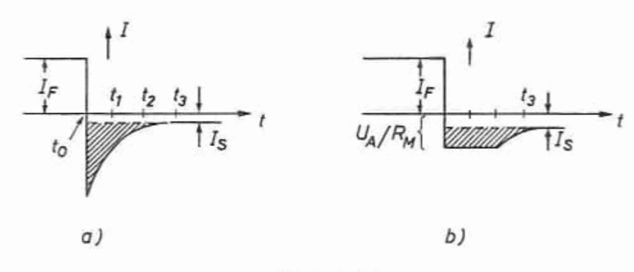


Abb. 5.46.
Zeitlicher Stromverlauf eines Sperrimpulses.

Frequenz angelegt und der Strom durch die Diode oszillographiert. Bei niedriger Frequenz (100 Hz) tritt das ein, was man auf Grund der statischen Strom-Spannungs-Kennlinie erwartet (Abb. 5.43): die negative Halbwelle des Wechselstromes wird weitgehend unterdrückt (Abb. 5.44a). Bei höheren Frequenzen

erscheinen jedoch zu Beginn der negativen Halbwelle kleine Stromspitzen, die mit steigender Frequenz größer werden (Abb. 5.44 b und c), bis bei höchsten Frequenzen die Diode ihre Gleichrichtereigenschaft mehr oder weniger verliert (Abb. 5.44 d).

5.6.1 Impulsverhalten. Diese Erscheinung läßt sich leichter übersehen, wenn die Diode mit Rechteckimpulsen belastet wird. In Abb. 5.45 ist die dazu verwendete Schaltung angegeben [52]. In der Periode zwischen den Impulsen ist die Germaniumdiode durch die Gleichstromquelle B in Flußrichtung belastet. Der Meßwiderstand R_M ist in diesem Zustand durch die Röhrendiode kurzgeschlossen, so daß nur eine kleine positive Spannung am Verstärker liegt.

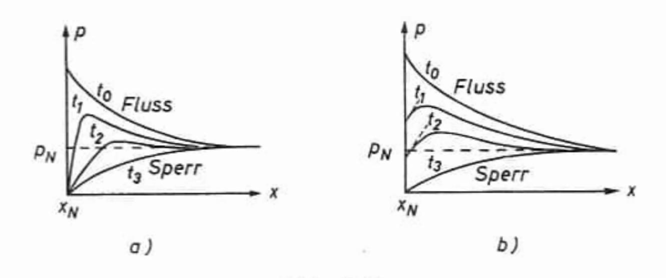


Abb. 5.47.
Minoritätendichteverteilungen zu verschiedenen Zeiten nach Einschalten des Sperrimpulses.

Beim Einsetzen des Impulses überlagert sich der von der Batterie stammenden Spannung an der Ge-Diode eine so hohe entgegengesetzte Spannung daß diese für die Dauer des Impulses in Sperrichtung vorgespannt wird. Gleichzeitig wird auch die Röhrendiode in Sperrichtung gepolt, so daß jetzt der Sperrstrom durch die Ge-Diode als Spannungsabfall an R_M auf den Verstärker gelangt und am Oszillographen sichtbar wird.

Das Oszillogramm dieses Stromes ist in Abb. 5.46 a schematisch dargestellt. Es ist dabei zu beachten, daß der Flußstrom wegen der Röhrendiode in sehr viel kleinerem Maßstab aufgezeichnet wird als der Sperrstrom. Wir sehen wieder, wie in Abb. 5.44, die Stromspitze zu Beginn der Sperrperiode. Die Verhältnisse sind aber jetzt relativ einfach zu übersehen. Während der Periode, in der der konstante Flußstrom fließt, stellt sich im Diffusionsgebiet eine Minoritätendichte $p(x)$ ein, die durch Gl. 5.60 mit $U > 0$ gegeben ist. Sie ist als strichpunktierte Linie in der logarithmischen Darstellung in Abb. 5.31 gezeichnet. Beim Anlegen des Sperrimpulses muß sich diese Verteilung in die neue Verteilung für $U < 0$ (Gl. 5.60), d. h. in die gestrichelte Verteilung von Abb. 5.31, umwandeln. Dies geschieht durch Diffusion der Minoritäten von der Diffusionszone zurück zum Rand der Feldzone, was sich in dem Stromstoß in Abb. 5.46a bemerkbarmacht. Eine ähnliche Kennlinie erhält man beim Ent-

laden eines Kondensators. Man spricht auch tatsächlich von einer „Diffusions"-Kapazität, die wir im nächsten Abschnitt noch genauer definieren werden.

Dieses Umladen der Diffusionszone ist in Abb. 5.47a dargestellt, wo die Minoritätenverteilungen in den verschiedenen Zeitpunkten der Abb. 4.46a aufgetragen sind. Man sieht, wie die Ladungsträgerdichte p am Feldzonenrand x_N sofort nach Anlegen der Sperrspannung gleich Null wird. Dadurch wird der Strom $j_p = - e\, D_p\, (\partial p / \partial x)_{x_N}$ sehr groß und nimmt langsam ab mit der Abnahme von $(\partial p/\partial x)_{x_N}$. Zur genauen Berechnung von $p(x, t)$ und $j_p(t)$ ist wieder eine Lösung der Bilanzgleichung 4.75 nötig, mit der Anfangsbedingung Gl. 5.60, mit $U > 0$ und mit den Randbedingungen $p(x_N) = 0$ und $p(\infty) = p_N$. Bezüglich der Durchführung dieser Rechnung muß hier jedoch auf die Literatur verwiesen werden [53, 54].

Eine Komplikation tritt auf, wenn der Strom, der zur Entladung der Diffusionskapazität fließen muß, durch den äußeren Widerstand R_M (Abb. 5.45) begrenzt ist. Dann ist der Strom durch $I \leq U_A/R_M$ gegeben, d. h. die Stromspitze ist abgeplattet (Abb. 5.46b). Entsprechend ist auch der Gradient $(\partial p/\partial x)_{x_N}$ der Minoritätendichte am Feldzonenrand begrenzt. Dies ist in Abb. 5.47b dargestellt. (Näheres darüber siehe SEILER und WUCHERER [52].)

Im Gegensatz zur Berechnung des Stromes $j_p(t)$ ist die Berechnung der Ladung der Diffusionskapazität, d. h. der schraffierten Fläche unter der Stromkurve in Abb. 5.46, sehr einfach. Es ist

$$Q = \int_{t_0}^{\infty} (I - I_s)\, dt = e\, A \int_{x_N}^{\infty} [p(U > 0) - p(U < 0)]\, dx$$

Darin ist für die Werte von p die Gl. 5.60 einzusetzen. A ist die Fläche der PN-Schicht. Man erhält

$$Q = e\, A\, p_N\, e^{U/U_T} \int_{x_N}^{\infty} e^{-(x - x_N)/L_p}\, dx = e\, A\, p_N\, e^{U/U_T}\, L_p$$

$$= \frac{e\, A\, D_p\, p_N}{L_p}\, e^{U/U_T}\, \tau_p = (I_F + I_S)\, \tau_p \approx I_F\, \tau_p \qquad \text{(Gl. 5.77)}$$

Hierbei ist stillschweigend vorausgesetzt worden, daß vernachlässigbar wenig Minoritätenträger während des Umschaltvorganges durch Rekombination verlorengehen. Unter dieser Voraussetzung ist also die Ladung Q der Diffusionskapazität gerade gleich der Ladungsmenge, die während der Flußperiode in einer Zeit τ_p in das Diffusionsgebiet hineingeflossen ist. Gl. 5.77 deutet auch an, daß diese Messung zur Bestimmung der Lebensdauer brauchbar ist, wenn der obenerwähnte Verlust durch Rekombination in Rechnung gestellt wird.

Ein ähnlicher kapazitiver Effekt tritt im Prinzip auch beim Umschalten von Sperr- auf Flußrichtung auf, da die Diffusionszone wieder mit Löchern angefüllt werden muß. Er wird jedoch häufig durch den Bahnwiderstand des Gleichrichters unterdrückt. Neben dem kapazitiven Effekt tritt beim Einschalten eines Flußimpulses noch ein induktiver Effekt auf, der durch die allmähliche Erhöhung der Leitfähigkeit im Halbleiterinnern auf Grund der Minoritäteninjektion zustande kommt. Je nach der Größe der verschiedenen Parameter kann der kapazitive oder induktive Effekt überwiegen, oder es tritt sogar beim Einschalten des Flußimpulses eine kurze Oszillation des Stromes auf. (Näheres siehe z. B. Wen-Hsiung Ko [55].)

5.6.2 Wechselstromverhalten. Nach den Ausführungen im vorhergehenden Abschnitt ist einzusehen, daß der Strom bei großen Wechselspannungsamplituden infolge der Trägheit der Diffusionszonen einen komplizierten zeitlichen Verlauf haben wird. Der Fall kleiner Spannungsamplituden dagegen ist verhältnismäßig übersichtlich. Da damit auch das Verständnis des Hochfrequenzverhaltens der Transistoren erleichtert wird, soll hier näher darauf eingegangen werden.

Wir nehmen also an, daß an der *PN*-Schicht eine Spannung

$$U = U_0 + u_1\, e^{i\omega t} \tag{Gl. 5.78}$$

liegt. u_1 sei $\ll U_T$, es handelt sich also um sehr kleine Amplituden der Wechselspannungskomponente.

Da die Feldzone den Spannungswechseln rasch zu folgen vermag, genügt es, nur das Verhalten der Minoritäten in der Diffusionszone zu betrachten. Es ist nach Gl. 5.34

$$p(x_N,t) = p_N\, e^{U/U_T} \approx p_N\, e^{U_0/U_T}\left(1 + \frac{u_1}{U_T}\, e^{i\omega t}\right) = p_N\, e^{U_0/U_T} +$$

$$+ p_N\, \frac{u_1}{U_T}\, e^{i\omega t}\, e^{U_0/U_T} = p_g(x_N) + \tilde{p}(x_N, t) \tag{Gl. 5.79}$$

Der erste Term dieser Gleichung p_g gibt mit der Bilanzgleichung den bekannten Dichteverlauf bei Gleichspannung. Wegen der Linearität der Bilanzgleichung 4.75 genügt es also, die Wirkung der kleinen Spannungsmodulation auf die Minoritätendichte zu betrachten. Dies geschieht durch Lösung von Gl. 4.75 mit $g_L = 0$ und der Randbedingung 5.79 ohne den ersten Term. Man hat also als Differentialgleichung:

$$\frac{\partial \tilde{p}}{\partial t} = -\frac{\tilde{p}}{\tau_p} + D_p\, \frac{\partial^2 \tilde{p}}{\partial x^2} \tag{Gl. 5.80}$$

Die am Feldzonenrand vorliegenden Löcherdichteschwankungen setzen sich in Form gedämpfter, ebener Wellen ins Kristallinnere fort, weshalb der Ansatz

$$\tilde{p} \sim e^{i\omega\left(t-\frac{x}{v}\right)} \qquad \text{(Gl. 5.81)}$$

am Platze ist. Dabei ist die Größe v die komplexe Fortpflanzungsgeschwindigkeit der Welle. Einsetzen in Gl. 5.80 liefert

$$i\omega = -\frac{1}{\tau_p} - D_p \frac{\omega^2}{v^2} \qquad \text{(Gl. 5.82)}$$

oder

$$\frac{1}{v} = (\pm)\frac{i\sqrt{1 + i\omega\tau_p}}{\omega L_p} \equiv \frac{1}{v_r} - \frac{i}{\omega L_{pw}} \qquad \text{(Gl. 5.83)}$$

v_r ist die reelle Fortpflanzungsgeschwindigkeit und L_{pw} hängt mit der Dämpfung der Welle zusammen, wie man durch Einsetzen in Gl. 5.81 sieht

$$\tilde{p} \sim e^{i\omega\left(t-\frac{x}{v_r}\right)} \cdot e^{-\frac{x}{L_{pw}}} \qquad \text{(Gl. 5.84)}$$

Mit der Randbedingung Gl. 5.79 ergibt sich damit

$$\tilde{p} = p_N \frac{u_1}{U_T} e^{U_0/U_T} e^{i\omega\left(t-\frac{x-x_N}{v_r}\right)} e^{-\frac{x-x_N}{L_{pw}}} \qquad \text{(Gl. 5.85)}$$

Die Stromdichte einer solchen Welle für $x = x_N$, also am Anfang der Diffusionszone, ist

$$\begin{aligned} j_p &= -e D_p \operatorname{grad} p\big|_{x=x_N} \\ &= e D_p p_N \frac{u_1}{U_T} e^{U_0/U_T}\left[\frac{1}{L_{pw}} + \frac{i\omega}{v_r}\right] e^{i\omega t} \end{aligned} \qquad \text{(Gl. 5.86)}$$

d. h. der Strom ist gegen die Spannung phasenverschoben. Einfache Beziehungen erhält man daraus und aus Gl. 5.83 für

$$\omega\tau_p \ll 1 \quad \text{oder} \quad \omega\tau_p \gg 1$$

5.6.2.1 Niedrige Frequenz: $\omega\tau_p \ll 1$. Da τ_p in der Größenordnung von μ sec liegt, gilt diese Näherung für Frequenzen unterhalb etwa 100 kHz. Aus Gl. 5.83 erhält man

$$\frac{1}{v_r} = \frac{1}{2}\frac{\tau_p}{L_p}; \qquad L_{pw} = L_p \qquad \text{(Gl. 5.87)}$$

Die Wellenlänge λ der p_w-Welle ist demnach

$$\lambda = \frac{v_r}{\omega/2\pi} = \frac{4\pi L_p}{\omega \tau_p} \gg L_p \qquad \text{(Gl. 5.88)}$$

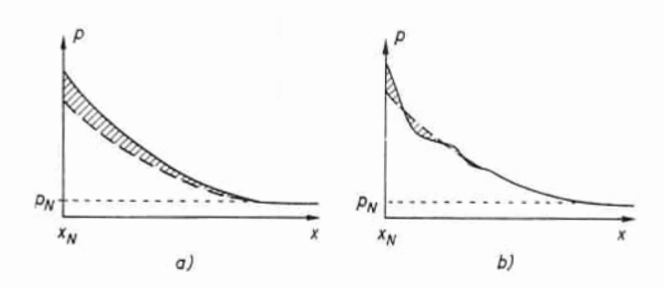


Abb. 5.48.
Minoritätendichteverteilung bei tiefer Frequenz (a) und hoher Frequenz (b). Gestrichelte Kurve: Gleichstromverteilung.

d. h. im Falle niedriger Frequenzen fällt die durch die Wechselspannung u_1 verursachte zusätzliche Löcherdichte ebenso exponentiell ab wie bei angelegter Gleichspannung (siehe Abb. 5.48a). Innerhalb dieses Bereiches folgt die Löcherdichteschwankung synchron der Wechselspannung mit einer kleinen Phasenverschiebung von der Größe $-\tau_p (x - x_N)/(2 L_p)$. Für die Wechselstromdichte erhält man nach Gl. 5.86 und 5.87

$$j_p = (G_p + i\omega C_{pD}) u_1 e^{i\omega t} \qquad \text{(Gl. 5.89)}$$

mit dem Wirkleitwert

$$G_p = \frac{e D_p p_N}{L_p U_T} e^{U_0/U_T} = \frac{j_{ps}}{U_T} e^{U_0/U_T} \qquad \text{(Gl. 5.90)}$$

und der Diffusionskapazität

$$C_{pD} = G_p \frac{\tau_p}{2} \qquad \text{(Gl. 5.91)}$$

Die Phasendrehung $\omega C_{pD}/G_p$ ist demnach gleich $\omega \tau_p/2$, wächst also mit der Frequenz.

Einen entsprechenden Ausdruck erhält man für die Elektronen der P-Diffusionszone. Wir wollen uns hier aber auf den Fall eines stark unsymmetrischen PN-Übergangs mit sehr niederohmigem P-Gebiet beschränken und können den Elektronenfluß ins P-Gebiet deshalb wegen der Kleinheit von n_P unberücksichtigt lassen.

Der Wirkleitwert kann nach Gl. 5.90 auch durch Differentiation der Gleichstromkennlinie im Arbeitspunkt U_0 gewonnen werden (vgl. Gl. 5.61). Der Blindleitwert ist unter der Voraussetzung $\omega\tau_p \ll 1$ immer klein gegenüber dem Wirkleitwert. Er kann durch eine Kapazität C_{pD}, die wir als Diffusionskapazität bezeichnet haben (auch SHOCKLEY-Kapazität genannt) und die zum Wirkleitwert parallel liegt, beschrieben werden. Diese Kapazität ist ebenso wie die in Abschnitt 5.4.2.4 beschriebene Feldzonenkapazität C_F (auch SCHOTTKY-Kapazität genannt) spannungsabhängig. C_F geht proportional $1/l$ und damit nach einer negativen Potenz mit der Spannung. C_D folgt dagegen exponentiell der Spannung. So kommt es, daß bei Sperrspannungen meist nur C_F wirksam ist und im Durchlaßgebiet nur C_D. Eine einfache Rechnung zeigt, daß bei Germanium für $U_0 = 0$ beide Kapazitäten von derselben Größenordnung sind. C_F ist für $U_0 = 0$ z. B. für den *PN*-Übergang von Abb. 5.22 nach Gl. 5.21 und 5.32 etwa von der Größe 18000 pF/cm². Für C_D erhält man bei Ge für $U_0 = 0$ und Zimmertemperatur mit $j_{ps} = 110\,\mu\text{A/cm}^2$ (aus Gl. 5.54) und $\tau_p = 3\,\mu\text{s}$ aus Gl. 5.90 und 5.91 einen Wert von 6600 pF/cm². Für $U_0 = 100$ mV ergibt sich für C_D bereits das Fünfzigfache!

5.6.2.2 Hohe Frequenzen: $\omega\tau_p \gg 1$. Aus Gl. 5.83 folgt näherungsweise für diesen Fall

$$\frac{1}{v_r} = \frac{1}{\omega L_{pw}} = \frac{1}{\sqrt{2}}\frac{1}{\sqrt{\omega D_p}} \qquad \text{(Gl. 5.92)}$$

und

$$L_{pw} = \sqrt{2}\sqrt{\frac{D_p}{\omega}} = \sqrt{2}\sqrt{\frac{D_p\tau_p}{\omega\tau_p}} \ll L_p \qquad \text{(Gl. 5.93)}$$

Die Löcherdichtewelle dringt also jetzt wesentlich weniger tief ein als bei kleineren Frequenzen. Die Eindringtiefe ist außerdem unabhängig von der Löcherlebensdauer τ_p. Die Wellenlänge der Löcherwelle ist nach Gl. 5.92

$$\lambda = \frac{v_r}{\omega/2\pi} = 2\pi L_{pw} \qquad \text{(Gl. 5.94)}$$

Die Eindringtiefe des gedämpften Wellenzuges ist also von der Größenordnung einer Wellenlänge (vgl. Abb. 5.48b). Die Dämpfung der Löcherdichtewelle hat jetzt nicht mehr die begrenzte Löcherlebensdauer als Ursache, sondern die Phasenunterschiede, die durch die relativ langsame Diffusion entstehen. Die Eindringgeschwindigkeit ist jedoch viel größer als im niederfrequenten Fall, wie man aus den Gleichungen 5.87, 5.92 und 5.93 leicht sieht.

Für $\omega\tau_p \gg 1$ wird das Glied $\tilde{p}/\tau_p$ in Gl. 5.80 vernachlässigbar. Dann hat Gl. 5.80 die Form der Wärmeleitungsgleichung. Diese Gleichung bestimmt z. B. das Eindringen der täglichen und jährlichen Temperaturschwankungen ins Erdinnere. Auch hier dringen die täglichen, also höherfrequenten Schwan-

kungen schneller, aber weniger tief ins Erdinnere als die niederfrequenten jährlichen. Die Dämpfung dieser Wellen kommt dadurch zustande, daß z. B. für negative Amplituden aus dem Erdinneren wieder Wärme zurückströmt. Dasselbe passiert mit den Löchern in der N-Diffusionszone.

Die Wechselstromwirk- und -blindleitwerte ergeben sich für $\omega\tau_p \gg 1$ aus Gl. 5.86 mit Gl. 5.92 und 5.93

$$G_p = \omega C_{pD} = \frac{e D_p p_N}{L_p} \frac{e^{U_0/U_T}}{U_T} \sqrt{\frac{\omega\tau_p}{2}} = G_p(\omega \text{ klein}) \cdot \sqrt{\frac{\omega\tau_p}{2}} \qquad \text{(Gl. 5.95)}$$

Wirk- und Blindleitwert sind jetzt gleich und, wie man durch Vergleich mit Gl. 5.90 feststellt, viel größer als der Leitwert bei niedrigen Frequenzen.

In Gl. 5.95 tritt nun auch τ_p wieder auf. Das kommt jedoch nur von der speziellen mit Gl. 5.90 vergleichbaren Form, die außerdem L_p und damit $\sqrt{\tau_p}$ auch im Nenner enthält. Man kann statt Gl. 5.95 auch schreiben

$$G_p = \omega\, C_{pD} = \frac{e\, p_N}{U_T}\, e^{U_0/U_T} \sqrt{\frac{\omega\, D_p}{2}} \qquad \text{(Gl. 5.96)}$$

Hier ist jetzt τ_p verschwunden, wie man erwarten muß.

5.6.2.3 Allgemeinfall. Durch Einsetzen von Gl. 5.83 in Gl. 5.86 erhält man den für *alle* Frequenzen gültigen Ausdruck

$$\begin{aligned} j_p &= (G_p + i\omega\, C_{pD})\, u_1\, e^{i\omega t} \\ &= \frac{e\, D_p p_N}{L_p} \frac{e^{U_0/U_T}}{U_T} \sqrt{1 + i\omega\tau_p}\, u_1\, e^{i\omega t} \end{aligned} \qquad \text{(Gl. 5.97)}$$

aus dem sich die Wirk- und Blindleitwerte für die obigen Grenzfälle unmittelbar ablesen lassen. Sie sind in Abb. 5.49 in doppeltlogarithmischem Maßstab gegen die Frequenz aufgetragen [34]. Dabei sind die theoretischen Kurven auf beiden Achsen so verschoben worden, daß sich optimale Anpassung an gemessene Punkte ergibt. Daraus erhält man einen Wert für τ_p und einen Wert für G_p bei $\omega = 0$. Insbesondere liefert der Schnittpunkt der drei Asymptoten den Frequenzwert, wo $\omega = 2/\tau_p$ ist. Dieses Verfahren kann also zur Lebensdauermessung herangezogen werden und ist insbesondere bei der Messung sehr kurzer Lebensdauern, wo die anderen Methoden versagen, von Nutzen.

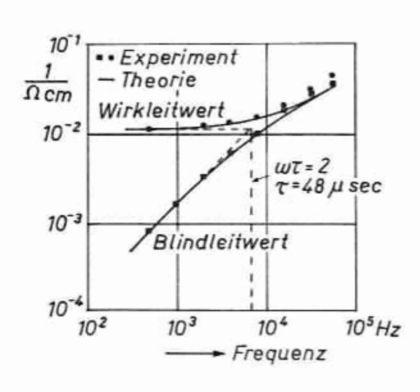


Abb. 5.49.
Wirkleitwert und Blindwert eines PN-Übergangs als Funktion der Frequenz [34]

5.7 Die Tunneldiode

5.7.1 Einführung. Im Jahre 1957 untersuchte der Japaner Esaki *PN*-Übergänge in sehr hoch dotiertem Germanium und entdeckte eine Kennlinie, die allem bisher bekannten radikal widersprach [56]. An Stelle der Kennlinienform der Abb. 5.7 oder 5.30 mit einem in Flußrichtung stark ansteigenden Strom fand er die Charakteristik von Abb. 5.9. Hier geht der Flußstrom mit steigender Spannung bei (etwa 60 mV) durch ein Maximum, sinkt dann mit weiter steigender Spannung, geht durch ein Minimum und steigt erst oberhalb 300 mV exponentiell an. In Sperrichtung steigt der Strom sehr steil an, so daß die Diode hier niederohmiger ist als in Flußrichtung.

Diese Entdeckung hat sofort außerordentliches Interesse in der Halbleiterindustrie und in der Halbleiterforschung hervorgerufen. Der Grund dafür ist, daß der Kennlinienteil mit dem negativen Anstieg einen negativen Widerstand für Wechselströme darstellt. Wie wir später sehen werden, kommt dieser negative Widerstand zustande durch den quantenmechanischen Tunneleffekt und hat nichts mit Minoritäteninjektion zu tun. Aus diesem Grunde arbeitet die Tunneldiode nahezu trägheitslos und sehr rauscharm. Sie ist geeignet zur Verstärkung von höchsten Frequenzen, wie sie mit Transistoren nicht mehr erreicht werden. Außerdem kann sie wegen der besonderen Kennlinienform (Abb. 5.9), die zudem nur wenig von der Temperatur abhängt, als Schalter verwendet werden. Andererseits ist der Effekt rein physikalisch sehr interessant, weil er weitere Einblicke unter anderem in die Bandstruktur von Halbleitern, insbesondere von entarteten Halbleitern, ermöglicht. Seither sind zahlreiche Arbeiten über Tunneldioden erschienen, von denen nur einige wenige im Literaturverzeichnis am Schluß des Buches angegeben werden können. Der Esaki-Effekt wurde nicht nur an Ge [57], sondern auch an Si [58], GaAs [59], InSb [60], und InAs [61] untersucht. Es wurden dabei grundsätzlich neue physikalische Phänomene entdeckt [62, 63, 64]. Wichtig sind auch die neuen Gesichtspunkte in der Schaltungstechnik, die sich bei der Anwendung der Tunneldiode, ergeben 65, 66, 67, 68, 69], ebenso wie neue Möglichkeiten von Tunnelkombinations-Bauelementen [70]. Eine gute Übersicht mit weiterer Literatur bietet die Arbeit von Hartmann et al. [71].

5.7.2 Der Zener-Effekt. Zum Verständnis der Tunneldioden soll zuerst der Durchbruchmechanismus behandelt werden, wie er in mäßig hochdotierten *PN*-Übergängen bei höheren Sperrspannungen auftritt und wie er zuerst von C. Zener [72, 73] berechnet wurde. Dazu gehen wir am besten von der Darstellung der Potentiale in der Sperrschicht in der Form des Bändermodells (Abb. 5.25) aus, das wir in Abb. 5.50 für eine angelegte Sperrspannung zeichnen. Dadurch wird die *N*-Seite auf eine tiefere Energie abgesenkt, während die ge-

samte Spannung in der Raumladungszone abfällt (vgl. Abschnitt 5.4.2.3). Die Feldstärke $F_{\max}$ in der Sperrschicht, (die Neigung der Bandkanten in Abb. 5.50), nimmt nach Gl. 5.31 mit der Spannung zu. Damit nimmt der waagrechte Abstand w zwischen den beiden Bändern ab nach der Gleichung (siehe Abb 5.50):

$$w = \frac{E_G}{e F_{\max}} \approx \frac{E_G l_0}{2e \sqrt{V_D |U|}} \qquad \text{(Gl. 5.98)}$$

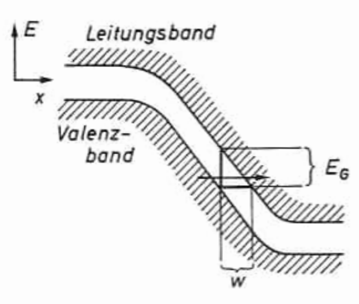


Abb. 5.50.
Zum ZENER-Effekt.

Bei einer gewissen kritischen Sperrspannung ist w so klein, daß es den Elektronen möglich ist, die Bandlücke waagrecht unter Erhaltung der Energie zu durchtunneln. Auf der linken Seite werden für diesen Tunnelprozeß sehr viele Valenzelektronen angeboten. Auf der rechten Seite werden die ankommenden Elektronen im fast leeren Leitungsband durch das hohe Feld nach rechts weitergeführt, so daß der Tunnelstrom ausschließlich von links nach rechts verläuft. Die Tunnelwahrscheinlichkeit Z hängt exponentiell von der Dicke der Potentialschwelle w, der Höhe E_G derselben und der effektiven Trägermasse m_{eff} ab:

$$Z \sim e^{-\text{const} \cdot w \cdot \sqrt{m_{\text{eff}} E_G}} \qquad \text{(Gl. 5.99)}$$

So kommt es, daß bei der oben erwähnten kritischen Spannung der Sperrstrom sehr stark zunimmt. Dieser ZENER-Durchbrucheffekt wird, wie in Abschnitt 5.2.3 ausgeführt, in den sogenannten ZENER-Dioden ausgenützt (siehe auch Abb. 5.7).

Es muß hier aber betont werden, daß im Durchbruchsgebiet hochsperrender Dioden, d. h. sehr schwach dotierter PN-Übergänge, ein anderer Mechanismus den ZENER-Effekt überwiegt. Es ist dies der sogenannte Lawinenprozeß, bei dem die im Feld beschleunigten Träger durch Stoß Trägerpaare erzeugen (also Elektronen aus ihrer Valenzbindung herausschlagen), die ihrerseits wieder weitere Paare erzeugen, usw. Wir wollen jedoch diesen letzteren Prozeß hier nicht weiterverfolgen, sondern lediglich auf die Literatur verweisen [74, 75].

5.7.3 Der ESAKI-Effekt. Die in Gl. 5.30 vorkommende Sperrschichtdicke bei Vorspannung Null, l_0, hängt von der Dotierung N_D und N_A beiderseits des PN-Überganges ab. Dementsprechend wird die kritische Tunneldicke w bei höherer Dotierung schon bei kleinerer Sperrspannung U erreicht. Dies ist in Abb. 5.51 schematisch dargestellt. Es sind die Kennlinien von vier verschiedenen Dioden gezeichnet, deren Dotierung mit steigender Numerierung zu-

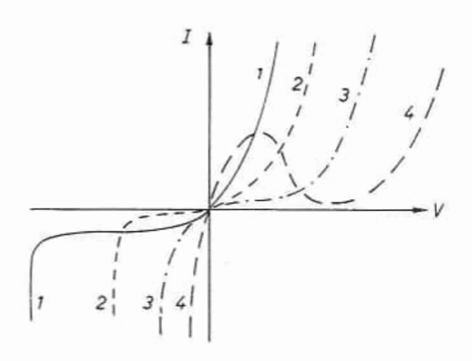


Abb. 5.51.
Übergang vom ZENER- zum ESAKI-Effekt.

nimmt. Kurve 1 habe beispielsweise N_D- und N_A-Werte von der Größenordnung 10^{16} cm^{-3}, während Kurve 2 etwa 10^{17} cm^{-3} entspricht und dementsprechend eine niedrigere Abbruchspannung hat. Bei Kurve 3 ($N_D \approx N_A \approx 10^{18}$ cm^{-3}) erfolgt der *Tunneldurchbruch* schon bei $U \approx 0$. Bei dieser Diode ist die „Sperrseite" niederohmiger als die Flußseite. Man nennt sie „verkehrter Gleichrichter". Solche Dioden haben Anwendungsmöglichkeiten in der Höchstfrequenztechnik, weil ihre Nullpunktskrümmung sehr viel größer ist als diejenige von normalen Dioden.

Bei weiterer Steigerung der Dotierung ($\approx 10^{19}$ cm^{-3}) entsteht schließlich die ESAKI-Kennlinie, und es liegt die Vermutung nahe, daß der hier auftretende Höcker in Flußrichtung ein ZENER-Tunnelstrom ist, der jetzt wegen der sehr hohen Dotierung sogar schon bei $U > 0$ in Erscheinung tritt. Außer der Verkleinerung der Sperrschichtdicke l_0 tritt bei der sehr hohen Dotierung noch

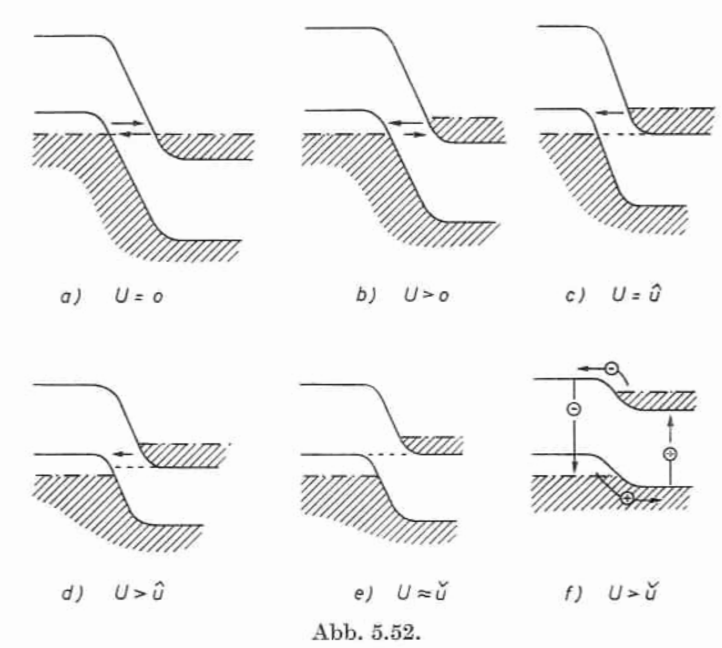


Abb. 5.52.
Bänderschema einer Tunneldiode bei verschiedenen Vorspannungen.

etwas Neues auf, nämlich die in Abschnitt 4.1.3 besprochene Entartung. In diesem Fall liegt die FERMI-Grenze nicht mehr in der Bandlücke, sondern im Valenz- bzw. Leitungsband. Diese Situation ist in Abb. 5.52a für $U = 0$ dargestellt. Zwischen dem teilweise leeren Valenzband und dem teilweise vollen Leitungsband fließen in beiden Richtungen Tunnelströme, die sich gegenseitig aufheben und $I = 0$ ergeben.

Wird eine kleine Spannung in Flußrichtung angelegt (Abb. 5.52b), so wird die Energie der Elektronen der rechten Seite (N-Seite) etwas erhöht. Es liegen dann den vollen Zuständen (schraffiert gezeichnet) des Leitungsbandes mehr leere Zustände des Valenzbandes gegenüber. Der Elektronentunnelstrom von rechts nach links ist also größer geworden. Umgekehrt liegen den vollen Zuständen des Valenzbandes weniger leere Zustände im Leitungsband gegenüber, so daß der Strom von links nach rechts kleiner geworden ist. Als Differenz ergibt sich ein Strom $I > 0$. Bei weiterer Steigerung der Flußspannung tritt schließlich der Fall ein, daß den vollen Zuständen des Valenzbandes gar keine Zustände des Leitungsbandes mehr gegenüberliegen (Abb. 5.52c). Dann bleibt nur noch ein Elektronenstrom von rechts nach links übrig. Bei dieser Spannung erreicht der Strom ein Maximum (Abb. 5.53). Weiteres Wachsen der Spannung führt zu einer Verringerung dieses Stromes, da jetzt weniger leere Plätze des Valenzbandes den vollen Zuständen des Leitungsbandes gegenüberstehen (Abb. 5.52d). Das ist das Gebiet der fallenden Charakteristik. Schließlich ist bei $U \approx \check{U}$ (Abb. 5.52e) gar kein Tunnelstrom mehr möglich. Der Strom geht durch ein Minimum. Erst bei höheren Spannungen setzt der normale Minoritätenstrom durch Injektion über den Potentialberg ein (Abb. 5.52f), was den bekannten exponentiellen Anstieg des Stromes mit der Spannung zur Folge hat (vgl. Gl. 5.62.)

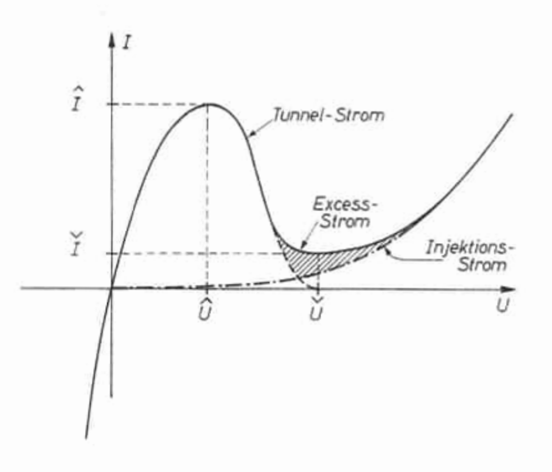


Abb. 5.53.
Beiträge zum Strom einer ESAKI-Diode

In Abb. 5.53 sind Tunnel- und Injektionsstrom getrennt eingezeichnet. Die experimentelle Kurve zeigt aber im Tal noch eine Stromkomponente, die mit der oben skizzierten Theorie nicht erklärt werden kann, und die „Excess“-Strom genannt wird. Er ist vermutlich auf Übergänge von und zu tiefliegenden Störstellenniveaus in der Bandlücke zurückzuführen [76].

Nach dem bisher Gesagten ist zu sehen, daß die genaue Form der ESAKI-Kurve von den Zustandsdichten an beiden Bandrändern und von der Tunnelwahrscheinlichkeit, d. h. vom genauen Tunnelmechanismus, bestimmt wird. Eine Analyse dieser Kurve kann deshalb dazu dienen, Informationen über diese Größen zu erhalten.

Insbesondere sehen wir, daß der Ast mit dem negativen Widerstand allein durch einen trägheitslosen Tunnelprozeß bestimmt ist. Infolgedessen ist die Frequenzgrenze einer Tunneldiode lediglich durch das RC-Glied aus Tunnelwiderstand und Sperrschichtkapazität bestimmt. Es konnten deshalb z. B. mit Tunneldioden aus Galliumarsenid Oszillationen bis zu 103 GHz erzeugt [69] und bei 26 GHz noch eine Verstärkung von 36 db erreicht werden [68].

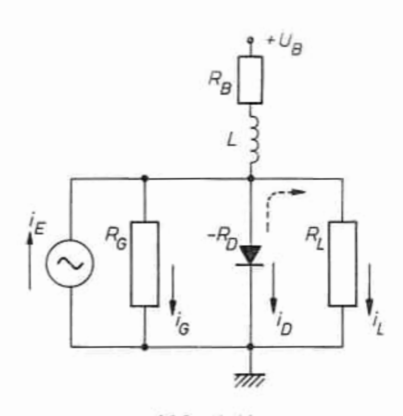


Abb. 5.54.
Einfacher Tunneldiodenverstärker.

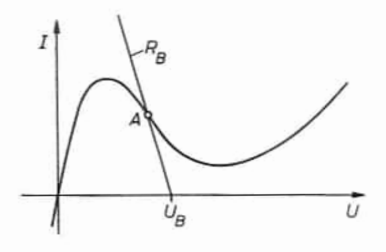


Abb. 5.55.
Arbeitspunkt des Tunneldiodenverstärkers.

5.7.4 Anwendung als Verstärker. Es ist angebracht, hier das Prinzip der Verstärkung an einem negativen Widerstand am Beispiel eines einfachen Tunneldiodenverstärkers zu erläutern, dessen Schaltbild in Abb. 5.54 angegeben ist. Die Diode ist über die Drossel L und den Widerstand R_B an die Gleichspannung $+ U_B$ angeschlossen. Damit ist der Arbeitspunkt A am fallenden Bereich der Charakteristik (Abb. 5.55) festgelegt. Die Widerstandsgerade darf aus Gründen der Stabilität die Diodenkennlinie nur einmal schneiden; es muß also $R_B < R_D$ sein, wenn $-R_D$ die Neigung der Diodenkennlinie bei A ist [66, 71]. Die Drossel L verhindert, daß die Diode über den Batteriekreis wechselstrommäßig kurzgeschlossen wird. R_G und R_L sind die Widerstände von Generator und Last. Für die Wechselströme gilt nach Abb. 5.54:

$$i_E = i_G + i_D + i_L \qquad \text{(Gl. 5.100)}$$

und

$$i_G R_G = i_L R_L = -i_D R_D \qquad \text{(Gl. 5.101)}$$

Daraus geht hervor, daß $i_D < 0$ ist, d. h. der Strom i_D fließt tatsächlich nicht in Richtung des ausgezogenen Pfeils in Abb. 5.54, sondern entgegengesetzt (gestrichelter Pfeil). Der Strom $|i_D| = - i_D$ stellt also einen zusätzlichen Strom dar, der durch die Last R_L fließt. Mit anderen Worten: eine Erhöhung der Spannung an der Diode ruft eine Stromänderung

entgegen der angelegten Spannung hervor, die dann als zusätzliche Stromerhöhung in dem parallel liegenden Lastwiderstand auftritt. Das stellt eine Stromverstärkung dar.

Aus Gl. 5.100 und 5.101 erhält man

$$i_L = \frac{G_L}{G_L + R_g - G_D} i_E \qquad \text{(Gl. 5.102)}$$

wo die *G's* die entsprechenden Leitwerte sind. Daraus geht die verstärkende Wirkung von $-G_D$ unmittelbar hervor.

Bei höheren Frequenzen entsteht jedoch durch die Sperrschichtkapazität der Diode C ein Nebenschluß, der eine Neutralisation durch eine Induktivität nötig macht. Dies ist in der Schaltung Abb. 5.56 durchgeführt. Dabei ist die Induktivität L gleichzeitig als Drossel für den Batteriekreis verwendet. C_B ist eine sehr große Kapazität. Dann ist G_D in Gl. 5.102 durch

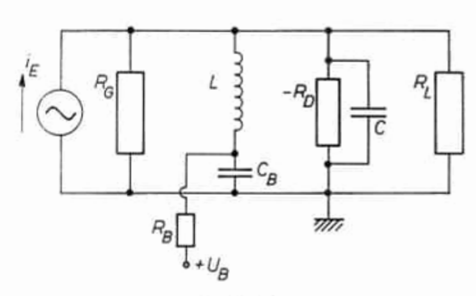


Abb. 5.56.
Tunneldiodenverstärker mit Neutralisation.

$$G_D + i\left(\omega C - \frac{1}{\omega L}\right) \text{ zu ersetzen. Damit wird}$$

$$i_L = \frac{G_L\, i_E}{G_L + G_G - G_D - i\,(\omega C - 1/\omega L)} \qquad \text{(Gl. 5.103)}$$

Die maximal vom Generator direkt verfügbare Leistung ist $i_E^2\, R_G/4$, die Nutzleistung am Ausgang ist $i_L\, i_L^*\, R_L$. Somit ist die Leistungsverstärkung

$$V_L = \frac{4G_G\, G_L}{(G_G + G_L - G_D)^2 + (\omega C - 1/\omega L)^2} \qquad \text{(Gl. 5.104)}$$

Vollkommene Neutralisation wird nur erreicht bei der Resonanzfrequenz

$$\omega_0 = \frac{1}{\sqrt{LC}} \qquad \text{(Gl. 5.105)}$$

Bei jeder Frequenz $\omega \neq \omega_0$ wird die Leistungsverstärkung verringert durch den Beitrag des zweiten Gliedes im Nenner von Gl. 5.104. Daraus ergibt sich eine endliche Bandbreite, die nur von dem Glied $R_D C$ abhängt.

Es ergibt sich nämlich nach Gl. 5.104 und 5.105 für diejenige Frequenzabweichung, für die V_L auf die Hälfte abgefallen ist,

$$\omega - \omega_0 \approx \frac{G_G + G_L - G_D}{2\,C} \qquad \text{(Gl. 5.106)}$$

Das Produkt aus Bandbreite $\Delta f = 2(\omega - \omega_0)/2\pi$ und der Wurzel aus der Leistungsverstärkung bei $\omega = \omega_0$ ist dann (als sogenannter Gütefaktor):

$$GB \equiv V_L^{1/2}\,\Delta f = \frac{2\sqrt{G_L G_G}}{2\,\pi\,C} \qquad \text{(Gl. 5.107)}$$

Für den Fall der optimalen Anpassung ($G_L = G_G$ und $G_D = G_L + G_G$) erhält man

$$GB = V_L^{1/2}\,\Delta f = \frac{1}{2\,\pi\,R_D\,C} \qquad \text{(Gl. 5.108)}$$

Man sieht also, daß bei gleichbleibender Kapazität die Tunneldiode als Hochfrequenzverstärker um so besser geeignet ist, je kleiner R_D, d. h. je steiler die I-U-Kennlinie im negativen Teil ist. Da dieses R_D sehr stark mit steigender Dotierung abfällt, während C nur schwach von der Dotierung abhängt, ist man bestrebt, solche Hochfrequenztunneldioden aus möglichst hochdotiertem Material herzustellen.

5.7.5 Anwendung als Schalter. Wird der Widerstand R_B, der im Gleichstromkreis in Serie mit der Diode liegt, im Gegensatz zum oben Gesagten größer als R_D gemacht, so wird die ESAKI-Kurve von der Widerstandsgeraden dreimal geschnitten (Abb. 5.57), wobei der Punkt A jetzt unstabil ist, wie aus einer genaueren Analyse unter Berücksichtigung der Diodenkapazität hervorgeht. Die Arbeitspunkte B und C sind stabil.

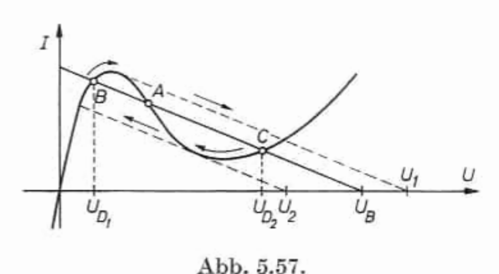


Abb. 5.57.
Arbeitspunkte des Tunneldiodenschalters.

Wird die Batteriespannung von $U = 0$ auf U_B gebracht, so stellt sich der Arbeitspunkt B ein, in dem über der Diode die relativ kleine Spannung U_{D_1} liegt. Vergrößert man aber die Batteriespannung kurzzeitig auf U_1 oder mehr, dann springt das System entlang den Pfeilen in Abb. 5.57 nach höheren Diodenspannungen und bleibt nach Beendigung des Zusatzspannungsimpulses am Arbeitspunkt C (Spannung U_{D_2}) stehen. Es kann durch kurzzeitiges Senken der Batteriespannung unter U_2 wieder in den Punkt B zurückgeschaltet werden. Man kann also durch geeignete Spannungsimpulse

die Diode vom niederohmigen in den hochohmigen Zustand und wieder zurückschalten. Diese Impulse können auch direkt an die Diode angelegt werden, wie dies in Abb. 5.58 geschehen ist.

Solche Schaltungen werden in Elektronenrechnern verwendet. Das System hat ein Gedächtnis, d. h. es zeigt am Ausgang $U_D = U_{D1}$ oder $U_D = U_{D2}$, je nachdem es irgendwann seit dem Einschalten einen positiven Spannungsimpuls von geeigneter Größe und Richtung erhalten hat. Man kann auch, wie in Abb. 5.58, zwei Eingänge (oder mehr) anbringen. Sind die Impulse so bemessen, daß ein einziger Impuls auf Klemme „1“ oder „2“ genügt zum Durchschalten der Diode, dann haben wir ein sogenanntes „Oder-Gatter“. Ist die Verabreichung beider Impulse auf „1“ und „2“ gleichzeitig notwendig, dann hat man ein „Und-Gatter“.

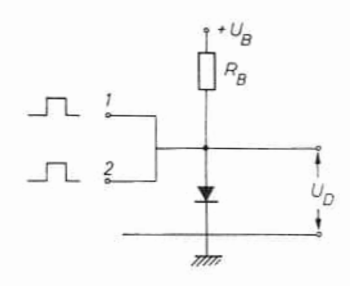


Abb. 5.58.
Tunneldiode als Schalter.

Für solche logischen Schaltaufgaben in Rechenmaschinen ist es besonders wichtig, daß das Umschalten schnell erfolgt. Man kann zeigen, daß, wie für die Verstärkeranwendung, auch hier die Größe $R_D C$ entscheidend ist und deshalb klein gemacht werden muß.

Sechstes Kapitel

Der Transistor

6.1 Einleitung

Die Entdeckung bzw. Erfindung des Transistors durch BARDEEN, BRATTAIN [37] und SHOCKLEY [33] im Jahre 1948 war in zweierlei Hinsicht von großer Bedeutung. Erstens hat der Transistor als Bauelement revolutionierend auf die Technik der Elektronik gewirkt; wir werden auf seine Vorzüge gegenüber der Elektronenröhre im nächsten Abschnitt (6.2) zu sprechen kommen. Zweitens aber hat dieses Ereignis außerordentlich befruchtend und anregend auf die gesamte festkörperphysikalische Forschung gewirkt, die seit dem Jahre 1948 fast lawinenartig angewachsen ist. Die Ursachen für diese Entwicklung liegen wohl darin, daß der Transistor, der auf den ersten Blick durch seine Einfachheit überrascht und deshalb technologisch attraktiv erscheint, im Grunde genommen ein Bauelement ist, dessen Verständnis und technische Beherrschung eine große Menge Grundlagenkenntnisse erfordert. Aus diesem Grunde steht das Transistorkapitel auch zu Recht am Schluß dieses Buches.

Der Aufbau und die Herstellung verschiedener Transistortypen ist in den Abschnitten 5.2.6 und 5.3 schon kurz besprochen worden. Wir wollen hier zunächst in vereinfachter und qualitativer Behandlung die Wirkungsweise des Transistors erläutern und erst in späteren Abschnitten eine saubere quanti-

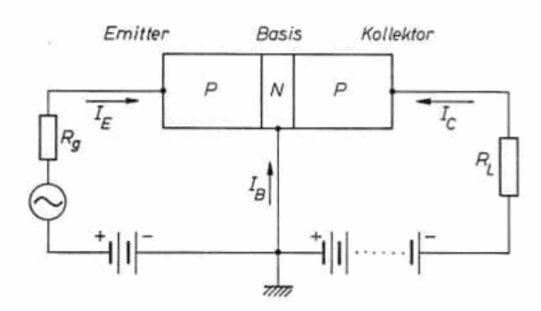


Abb. 6.1.
Prinzipielle Schaltung des Transistors.

tative Lösung des Problems angehen. Dabei ist jedoch auch dort die Behandlung nicht erschöpfend und nur die wichtigsten grundsätzlichen Zusammenhänge können aufgezeigt werden. Sowohl den weiter interessierten Transistortechnologen als auch den Schaltungstechniker müssen wir darüber hinaus auf die einschlägige Literatur verweisen, z. B. auf die im Literaturverzeichnis angeführten Bücher.

Abb. 6.1 zeigt das Schema eines *PNP*-Transistors in einer einfachen Verstärkerschaltung mit einem Generator am Eingang (Generatorwiderstand R_g) und einer Last R_L am Ausgang. Wie schon in Abschnitt 5.2.6 erwähnt, besteht der Transistor aus zwei *PN*-Übergängen, die so eng benachbart sind, daß eine Wechselwirkung zwischen den Vorgängen an diesen beiden *PN*-Schichten auftritt. Das ist nach den Ausführungen im letzten Kapitel nur möglich, wenn der Abstand der beiden *PN*-Übergänge, also die Dicke der Basisschicht, kleiner als die Minoritäten-Diffusionslänge ist, da sich nur dann die Diffusionszonen beider Übergänge überlappen. Das sind für Löcher in Germanium (nach Gl. 5.51 und Tab. 4) mit $\tau_p \approx 100\,\mu$sec etwa $50\,\mu$. Die Basisschicht muß also sehr dünn sein. Diese Forderung wird für Hochfrequenztransistoren noch weiter verschärft, da dort die Laufzeiten der Minoritäten zwischen den beiden Sperrschichten eine entscheidende Rolle spielen. So benötigt man bei manchen Typen Basisdicken, die unter $1\,\mu$ liegen.

Wie in Abb. 6.1 gezeichnet, sind im Verstärkerbetrieb der Emitterübergang in Flußrichtung (*P*-Seite positiv) und der Kollektorübergang in Sperrrichtung (*P*-Seite negativ) vorgespannt. Die eingezeichneten Pfeilrichtungen der Ströme entsprechen nicht dem wirklichen Stromfluß sondern der kon-

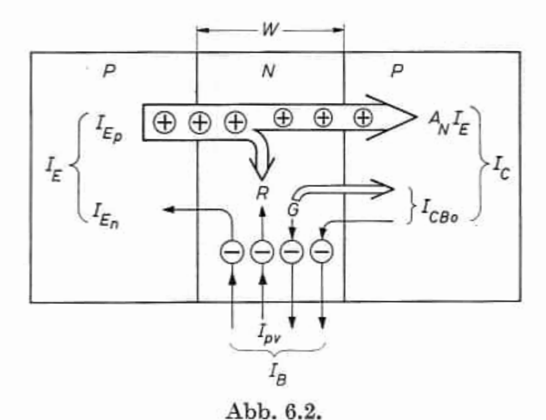


Abb. 6.2.
Die Stromkomponenten des Transistors.

ventionellen Richtungsfestlegung, wonach die Ströme in den Transistor hinein als positiv zu zählen sind. So ist beim *PNP*-Transistorverstärker $I_E > 0$, $I_C < 0$ und, wie wir sehen werden, I_B ebenfalls < 0. Beim *NPN*-Transistor ist es umgekehrt.

Wir wollen uns bei allen folgenden theoretischen Erörterungen der Einfachheit halber auf *PNP*-Transistoren beschränken. Das bedeutet aber in Wirklichkeit keinerlei Einschränkung, da alle Überlegungen und Formeln ohne weiteres auf *NPN*-Transistoren übertragen werden können, wenn jeweils der umgekehrten Polarität in geeigneter Weise Rechnung getragen wird.

Durch die in Flußrichtung vorgespannte Emittersperrschicht werden Löcher aus dem *P*-Gebiet in die *N*-Zone injiziert und diffundieren als Minoritäten durch die Basis hindurch zum Kollektor (siehe Abb. 6.2). Die am linken Rand der Kollektorfeldzone ankommenden Löcher werden von dort durch das hohe Feld in die Kollektor-*P*-Zone abgesaugt, da die am Kollektor liegende Sperrspannung ($U_{CB} < 0$) bestrebt ist, nach Gl. 5.34 die Löcherdichte am linken Rand der Kollektorfeldzone auf nahezu Null zu halten. Ein Teil der Löcher wird in der Basis rekombinieren bevor sie zum Kollektor gelangen und einen Elektronenstrom in die Basis hinein zur Folge haben. Dieser Teil ist aber klein, wenn die Basisdicke $W < L_p$ ist. Ein anderer Teil des Emitterstromes besteht aus einem Elektronenstrom von der Basis zum Emitter. Aber auch dieser Anteil kann klein gemacht werden durch eine hohe Dotierung der Emitterseite (vgl. Gl. 5.62, 5.63).

Durch die Kollektorsperrschicht fließt also erstens der normale Sättigungsstrom der Kollektordiode I_{CBO} (wobei $I_{CBO} < 0$ ist) und zweitens derjenige Bruchteil A_N des Emitterstromes I_E, der als Löcherstrom den Kollektorrand erreicht. Wegen der obigen Zeichenfestsetzung erhält man für den Kollektorstrom

$$I_C = I_{CBO} - A_N I_E \tag{Gl. 6.1}$$

Eine Änderung des Emitterstromes δI_E ruft also am Kollektor eine Änderung $\delta I_C = -A_N \delta I_E$ hervor. Wir nennen A_N in Übereinstimmung mit der üblichen Ausdrucksweise die „Stromverstärkung", obwohl sie immer < 1 ist. Sie ist allerdings bei guten Transistoren mit kleiner Basisdicke und hochdotiertem Emitter nur wenige Prozent kleiner als Eins.

Um den Emitterstrom um δI_E zu erhöhen, ist nur eine kleine Zusatzspannung δU_{EB} zwischen Emitter und Basis nötig, da der Emitterübergang in Flußrichtung gepolt und daher niederohmig ist. Der Kollektorübergang ist aber in Sperrichtung gepolt und daher hochohmig. Aus diesem Grunde ruft die Kollektorstromänderung $\delta I_C = -A_N \delta I_E \approx -\delta I_E$ eine viel größere Spannungsänderung am Kollektor hervor. Wir erhalten eine erhebliche Spannungsverstärkung und damit, weil die Stromverstärkung beinahe Eins ist, eine große Leistungsverstärkung.

Dieser Sachverhalt ist anschaulich an Hand von Abb. 6.3 einzusehen. In Abb. 6.3b ist der Kollektorstrom als Funktion der Kollektorspannung bei verschiedenen Emitterspannungen und damit auch Emitterströmen schematisch aufgetragen. (Die Kurven verlaufen in Wirklichkeit allerdings viel flacher.) Die Kurvenschar erinnert stark an die Kennlinie einer Photodiode

bei verschiedener Beleuchtung (Abb. 5.39). Auch dort wird ja der Strom durch zusätzliche Minoritätsträger, die am Rand der Sperrschicht angeboten werden, vergrößert. Abb. 6.3a zeigt die Eingangskennlinie, die im wesentlichen durch die Flußkennlinie der Emittersperrschicht gegeben ist.

Für $U_{EB} = U^{\circ}_{EB}$ folgt der Kollektorstrom der ausgezogenen Kennlinie in Abb. 6.3 b. Wird U_{EB} um eine kleine Spannung δU_{EB} vergrößert, so erhöht sich der Emitterstrom um δI_E, und es verschiebt sich die Kollektorkennlinie

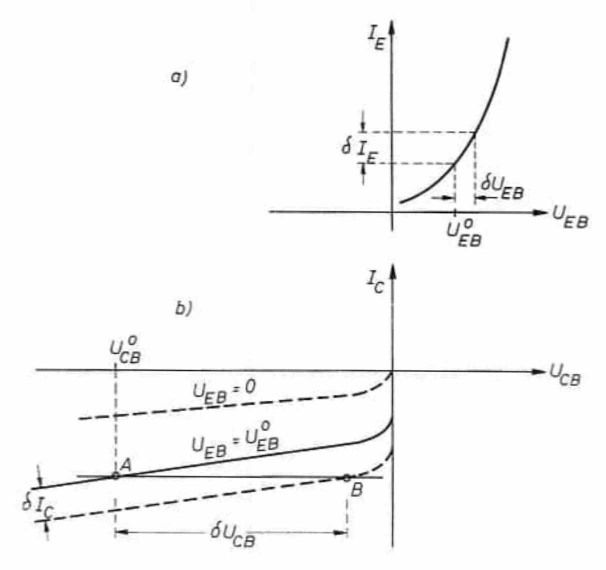


Abb. 6.3.
Zur Spannungsverstärkung: a) Eingangskennlinie, b) Ausgangskennlinien.

um $\delta I_C = -A_N \delta I_E \approx -\delta I_E$. Dies ergibt bei der Kollektorvorspannung U°_{CB} und einer sehr hochohmigen Last (Leerlauf) eine Veränderung δU_{CB} der Kollektorspannung von A nach B. Sind die Neigungen der Emitter- bzw. Kollektorkennlinien r_E bzw. r_C, so läßt sich aus Abb. 6.3 leicht ablesen, daß

$$\delta U_{CB} = r_C\, \delta I_C = -\, r_C A_N\, \delta I_E = -\, A_N \frac{r_C}{r_E}\, \delta U_{EB} \qquad \text{(Gl. 6.2)}$$

Wie schon aus der Abbildung hervorgeht, ist δU_{CB} ein Vielfaches von δU_{EB}. Wir haben eine große Leerlaufspannungsverstärkung. Typische Transistorkennlinien zeigen $r_E \approx 25\,\Omega$, $r_C \approx 120\,\text{k}\Omega$ und $A_N \approx 1$, was eine Leerlaufspannungsverstärkung $\delta U_{CB}/\delta U_{EB}$ von etwa 5000 ergibt. Die Leistungsverstärkung ist von der Größenordnung des Produkts aus Kurzschlußstromverstärkung A_N und Leerlaufspannungsverstärkung, genau genommen, wie wir weiter unten bei Gl. 6.104 sehen werden, gleich (1/4) $A_N(\delta U_{CB}/\delta U_{EB})$. Man

erhält also z. B. eine Leistungsverstärkung von 1000 oder, im logarithmischen Maß, von 30 dB.

6.2 Vergleich mit der Elektronenröhre

Nach dieser Einführung wollen wir, bevor wir zur strengeren Behandlung übergehen, ganz kurz die allgemeinen technischen Eigenschaften diskutieren, wie sie sich insbesondere aus einem Vergleich mit der Elektronenröhre ergeben.

6.2.1 Wirkungsgrad. Im Gegensatz zur Elektronenröhre, wo die Kathode auf Emissionstemperatur geheizt werden muß, arbeitet der Transistor bei Zimmertemperatur und ohne Heizkreis. Die für die Verstärkerwirkung nötigen Elektronen bzw. Defektelektronen sind im Halbleiter immer vorhanden. Sie müssen nicht, wie bei der Elektronenröhre, durch hohe Kathodentemperatur erst geschaffen werden. Der Transistorwirkungsgrad (Wechselstromausgangsleistung/Gleichstromverbrauch) ist deshalb denkbar günstig. So erreicht man z. B. im A-Betrieb fast 50%, im B-Betrieb etwa 70% Wirkungsgrad, was den theoretisch möglichen Werten sehr nahe kommt. Diese Tatsache bedeutet auch im anbrechenden Zeitalter der Energiefülle einen echten Fortschritt. Insbesondere ist sie von entscheidender Bedeutung für tragbare batteriebetriebene Geräte, wie z. B. Taschenradios, tragbare Sender, Hörhilfen usw. Sie ist auch von großer Wichtigkeit überall dort, wo sehr viele Bauelemente in dichter Packung zusammengebaut werden sollen, wie z. B. in elektronischen Rechenmaschinen, weil dann die niedrigere Wärmedissipation bei gleicher Leistung und gleicher zulässiger Temperatur im Gerät eine höhere Packungsdichte ermöglicht.

6.2.2 Spannungsbedarf. Ebenfalls für tragbare, batteriebetriebene Geräte ist der geringe Spannungsbedarf von großem Vorteil. Während Röhrengeräte Anodenspannungen von etwa 100 Volt benötigen, kommt ein Transistorgerät mit wenigen Volt aus.

Diese Tatsache kann z. B. mit Hilfe der in Abb. 6.4 skizzierten Anordnung demonstriert werden. Um den Transistoroszillator zum Schwingen zu bringen,

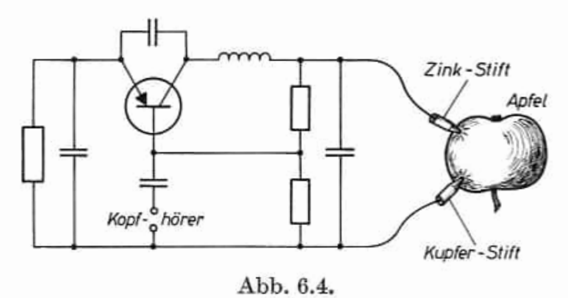


Abb. 6.4.
Demonstration des Spannungsbedarfs eines Transistoroszillators.

genügt als Stromquelle ein Kupfer- und ein Zinkdraht, die in eine Gurke oder einen Apfel gesteckt werden. Auch ein feuchtes Fließpapier oder Taschentuch zwischen beiden Drahtenden ist für kurze Zeit ausreichend. Die Schwingungen sind im Kopfhörer leicht nachzuweisen oder können mit Hilfe eines einfachen Transistorverstärkers vor einem größeren Kreis demonstriert werden.

6.2.3 Größe. Der Vorteil der Kleinheit des Transistors gegenüber der Elektronenröhre ist offensichtlich. Die Fortschritte in der Verkleinerung von elektronischen Geräten des täglichen Lebens durch die Verwendung von Transistoren sind bekannt. Es sei nur an die „Hörbrillen" erinnert, an Hörhilfen, die ganz in der Ohröffnung verschwinden, ferner an die Verkleinerung von Blitzlichtgeräten, Taschenradios und Polizei-Funkgeräten. Die Bedeutung dieser Entwicklung für militärische Zwecke bedarf keiner weiteren Erläuterung. Versuche mit Transistoroszillatoren, die zur elektrischen Stimulation eines kranken menschlichen Herzens in die Brusthöhle selbst einoperiert werden, beleuchten unter anderem die medizinischen Anwendungsmöglichkeiten, die dadurch erschlossen werden. Einer der größten Triumphe der Transistorelektronik sind vielleicht die Erfolge der Weltraumforschung mit Hilfe von Erdsatelliten und Raumschiffen, deren Meßergebnisse fast ausschließlich mit Hilfe von miniaturisierten Transistorgeräten aufgenommen, verarbeitet und zur Erde gesendet werden, wobei der Betriebsstrom vielfach von Silizium-Solarbatterien stammt.

Die Verkleinerung der Transistoren und Transistorschaltungen geht zur Zeit weiter in der Richtung, daß ganze Schaltungsteile mit Widerständen, Kapazitäten, Dioden und Transistoren auf einer Siliziumplatte erzeugt werden („Integrated Circuits", „Molecular Engineering"). Diese Entwicklung ist insbesondere für die elektronischen Rechenmaschinen von Bedeutung.

6.2.4 Betriebsbereitschaft. Ein Transistor ist unmittelbar nach dem Einschalten betriebsbereit. Im Gegensatz zu einer Röhre ist kein Aufheizen einer Glühkathode notwendig.

6.2.5 Lebensdauer. Da der Transistor keinen Heizfaden hat, fällt der wichtigste Faktor, der bei der Elektronenröhre die Lebensdauer begrenzt, weg. Die Lebensdauer von handelsüblichen Transistoren ist jedoch begrenzt durch Veränderungen, die an der Oberfläche des Germanium- oder Siliziumplättchens bezüglich ihres Belages mit Oxyden, Hydroxyden oder anderen Stoffen, die unter Umständen in monomolekularen Schichten schon von Einfluß sind, eintreten. Durch solche unvermeidlichen Fremdatomschichten, die sich an den freien Valenzen der Oberfläche anlagern, wird das elektrische Potential in einer oberflächlichen im Halbleiter liegenden Schicht beeinflußt. Dies wirkt sich auf die Größe der Leckströme und auf die Oberflächenrekombination aus und beeinflußt somit die Transistorkennlinie.

Dieses Problem ist in der Tat eines der wichtigsten und schwierigsten, mit denen der Transistorhersteller zu kämpfen hat. Der Aufwand, der zu diesem

Zweck getrieben wird, richtet sich selbstverständlich nach den Anforderungen, die vom Anwender gestellt werden. Durch sorgfältiges Ausheizen von Transistor und Gehäuse, durch Evakuieren des Gehäuses und Füllen mit sehr trockener Luft oder Stickstoff oder auch mit Luft von sehr genau kontrolliertem Feuchtigkeitsgehalt, durch Verwendung von Gettern usw. ist es gelungen, für die Telefonindustrie, insbesondere für Kabelverstärker, und für große Elektronenrechner Transistoren zu entwickeln, für die eine mittlere Lebensdauer von der Größenordnung 10^8 bis 10^9 Stunden, das sind 10^4 bis 10^5 Jahre [77], abgeschätzt werden kann.

6.2.6 Schockfestigkeit. Da der aktive Bereich eines Transistors sehr viel kleiner ist als der einer Röhre und sich hier das wesentliche Geschehen im Innern eines festen Körpers abspielt, kann der Transistor sehr viel kompakter aufgebaut werden. Die Schockfestigkeit ist deshalb schon bei normalen Transistortypen (wie z. B. dem in Abb. 5.17 dargestellten Typ) mit etwa 5000 g wesentlich besser als bei Röhren. Sie kann mit entsprechenden Kontaktierungs- und Einbaumethoden fast unbegrenzt gesteigert werden.

6.2.7 Schaltungsvereinfachung. Außer den Schaltungsvereinfachungen, die im Wegfall der Heizung oder dem geringen Spannungs- und Leistungsbedarf ihre Ursache haben, bieten komplementäre Elemente, z. B. *PNP*-Transistoren zusammen mit *NPN*-Transistoren, grundsätzliche Vereinfachungsmöglichkeiten. Meist fallen Übertragungsglieder fort. Da der komplementäre Transistor einer Röhre entspricht, deren Kathode Löcher emittiert, kann die Röhre diese Vereinfachung grundsätzlich nicht bieten. Auch die Anwendung des

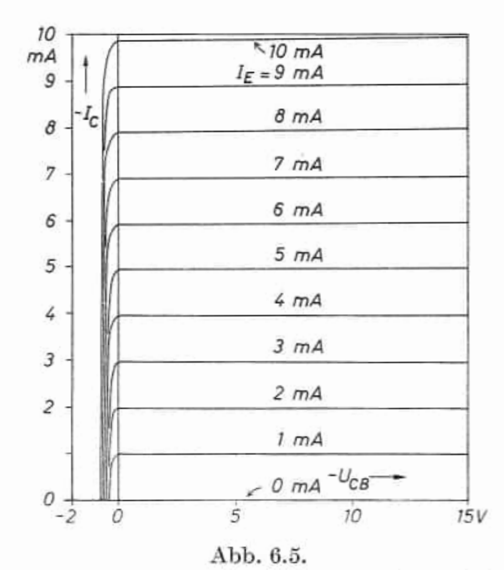


Abb. 6.5.
I_C-U_{CB}-Kennlinien des Standardtransistors OC 304.

Transistors als mono- oder bistabiler Schalter führt im allgemeinen auf einfachere Schaltungen als bei Röhren.

6.2.8 Kennlinienfeld. Abb. 6.5 zeigt die Kollektor-Kennlinie eines Transistors für verschiedene Emitter-Ströme. Aus dem Verlauf erkennt man die geringe Änderung der Stromverstärkung $\delta I_C/\delta I_E$ mit der Kollektorspannung und dem Emitterstrom. Die Verstärkung ist demnach nur mit geringen Verzerrungen behaftet. Diese Struktur des Kennlinienfeldes bleibt herab bis zur Spannung von einigen 0,1 Volt erhalten, d. h. eine verzerrungsfreie Verstärkung ist bis zu kleinsten Spannungen möglich. Pentoden haben ähnliche Kennlinienfelder, aber nur oberhalb von etwa 100 Volt. Ein Absinken der Betriebsspannung führt deshalb dort meist zu Verzerrungen.

6.2.9 Nachteile. Es sind im wesentlichen drei Richtungen, in denen der Anwendungsbereich der Transistoren gegenüber der Elektronenröhre begrenzt ist. Sie sind hohe Frequenz, hohe Ausgangsleistung und hohe Temperatur.

Die höchsten Frequenzen, die bisher von Transistoren erreicht worden sind, liegen bis 1–3 GHz, und zwar sind es hier im wesentlichen die Germanium-Mesatransistoren. Eine weitere Steigerung der Frequenzgrenze ist zum mindesten für serienproduzierte Transistoren nicht zu erwarten. Dagegen erreicht man bekanntlich mit Röhrendioden bis 4 GHz, mit Klystrons und Wanderfeldröhren bis 30 GHz und mit dem Magnetron annähernd 100 GHz.

Die größten Leistungstransistoren liegen bei Ausgangsleistungen von etwa 300 Watt. Obwohl es prinzipiell möglich ist, diese Zahl zu vergrößern, werden die höheren Leistungen wohl die Domäne der Röhren bleiben. Das Hauptproblem ist dabei, daß eine Vergrößerung der PN-Flächen zu einer rapiden Verschlechterung der Qualität und der Ausbeute führt auf Grund von Baufehlern im Kristall und von Unregelmäßigkeiten im Legier- bzw. Diffusionsprozeß.

Bei höherer Temperatur steigt die Eigenleitungsdichte an und verwischt schließlich den Unterschied zwischen P- und N-Material. Wir haben schon in Abschnitt 5.4.3.3 gesehen, daß aus diesem Grund der Sperrstrom mit der Temperatur stark ansteigt. Die Erhöhung des Kollektorsperrstromes mit der Temperatur bedingt wiederum erhöhte Verluste und weitere Erwärmung.

So kommt es, daß ein stabiler Betrieb von Germaniumtransistoren oberhalb 100° C nicht mehr möglich ist. Silizium kann wegen des größeren Bandabstandes noch bis etwa 250° C verwendet werden. Andere Halbleiter mit noch höherem Bandabstand, wie z. B. Galliumarsenid ($E_g = 1{,}35$ eV), versprechen in dieser Hinsicht mehr.

6.3 Gleichstromkennlinien bei kleinen Strömen

Wir wollen nun daran gehen, die Transistorkennlinien in strengerer Form abzuleiten, als dies im Einleitungsabschnitt geschehen ist. Dazu ist es jedoch nötig, den realen Transistor, der z. B. die Geometrie von Abb. 5.17b habe, zu

idealisieren, indem wir einen zylindrischen oder prismatischen *PNP*-Stab betrachten, wie er z. B. durch Ziehen aus der Schmelze und Umdotieren nach Abschnitt 5.3.1 und Abb. 5.14 oder durch Herausschneiden eines Stückes aus der Mitte der legierten Struktur von Abb. 5.17b entsteht. Außerdem soll die Oberflächenrekombination der Löcher an der Oberfläche der Basiszone vernachlässigt werden. Auf diese Weise erhalten wir ein lineares Problem, das den Randeinfluß und den Effekt eines Größenunterschiedes der Emitter- und Kollektorfläche vernachlässigt. Dies ist der idealisierte, innere Transistor, der in den folgenden beiden Abschnitten behandelt werden soll. Im Abschnitt 6.3.3 soll dann der sehr wichtige Einfluß des Querwiderstandes zwischen Basiskontakt und der Basisschicht des inneren Transistors eingeführt und ein Ersatzschaltbild des idealen Transistors angegeben werden. In dem darauffolgenden Abschnitt 6.3.4 wird dann auf die Abweichungen von dieser Idealisierung eingegangen.

6.3.1 Der idealisierte, innere Transistor. Wir wollen also zunächst das oben beschriebene Modell, eindimensionaler Fall ohne Oberflächenrekombination, für kleine Ströme berechnen und uns dabei auf Abb. 6.6 beziehen. Es müssen zunächst die Löcher- und Elektronendichten in den drei Zonen (Emitter $x < 0$, Basis $0 < x < W$, Kollektor $x > W$) als Funktionen der beiden Spannungen an den *PN*-Übergängen $U_{EB'}$ und $U_{CB'}$ berechnet werden.

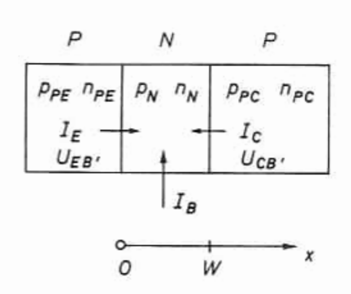


Abb. 6.6. Zur Berechnung des idealisierten Transistors.

In der Emitter- und der Kollektorzone sind die Löcher als Majoritäten wenig verändert. Wir können zum mindesten bei kleinen Strömen schreiben:

$$x < 0: \quad p = p_{PE} \qquad \text{(Gl. 6.3)}$$

und

$$x > W: \quad p = p_{PC} \qquad \text{(Gl. 6.4)}$$

Die Löcherdichte in der Basis ist $p(x) = p_N + \delta p$ wobei δp bestimmt wird durch die Bilanzgleichung 5.47 mit Gl. 5.51

$$\frac{\partial^2 \delta p}{\partial x^2} = \frac{\delta p}{L_p^2} \qquad \text{(Gl. 6.5)}$$

und durch die Randbedingungen (entsprechend Gl. 5.34 bzw. 5.59)

$$p(0) = p_N\, e^{\,U_{EB'}/U_T}\,; \quad \delta p(0) = p_N\left(e^{\,U_{EB'}/U_T} - 1\right) \qquad \text{(Gl. 6.6)}$$

und bei gesperrtem Kollektor, d. h. $U_{CB'} < 0$ und $|U_{CB'}| \gg U_T$

$$p(W) = p_N\, e^{U_{CB'}/U_T} \approx 0; \quad \delta p(W) = -p_N \qquad \text{(Gl. 6.7)}$$

Bevor wir an die exakte Lösung der Differentialgleichung 6.5 gehen, betrachten wir den sehr viel einfacheren und anschaulicheren Fall der verschwindenden Rekombination, also einen Transistor, dessen Basis so dünn ist, daß darin die Rekombination vernachlässigt werden kann. Dies gilt für $W \ll L_p$. Dann ist nämlich der Löcherstrom über der ganzen Basisdicke konstant und, da er ein Diffusionsstrom ist, gleich

$$I_{Ep} = e\, F\, D_p\, \partial p/\partial x = \text{const.}$$

Aus $(\partial p/\partial x) = \text{const.}$ folgt, daß die Löcherdichte $p(x)$ von $p(0)$ (Gl. 6.6) auf $p(W) = 0$ (Gl. 6.7) linear abfällt. Dies ist in Abb. 6.7 skizziert. Damit läßt sich der Ausdruck für den Strom sofort anschreiben

$$I_{Ep} = \frac{e\, F\, D_p\, p_N}{W}\, e^{U_{EB'}/U_T} \approx I_E \qquad \text{(Gl. 6.8a)}$$

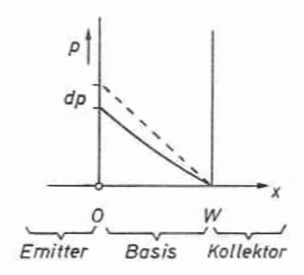


Abb. 6.7.
Die Minoritätendichte in der Basis.

Das letzte Glied der Gleichung gilt für den Fall, daß der Elektronenanteil des Emitterstromes $I_{En} \ll I_{Ep}$ ist, was wegen $n_{PE} \ll p_N$ fast immer zutrifft. Gl. 6.8a entspricht nach Gl. 5.62, 5.63 näherungsweise einer Diodenkennlinie, bei der die Diffusionslänge L_p durch die Basisdicke W begrenzt ist. Die Neigung dieser Kennlinie ist dann

$$r_E = 1 \Big/ \left(\frac{\partial I_{Ep}}{\partial U_{EB'}}\right) = \frac{U_T}{I_{Ep}} \approx \frac{U_T}{I_E} \qquad \text{(Gl. 6.8b)}$$

Für $I_E = 1$ mA erhält man (mit $U_T = 25$ mV) $r_E = 25\ \Omega$, was in Abschnitt 6.1 schon verwendet wurde.

Weiterhin können wir auch jetzt schon eine Abschätzung machen über den Teil I_{pV} des Emitterstromes, der durch Rekombination in der Basis verloren geht bzw. als Basisstrom abgeführt wird (Abb. 6.2). Er ist pro Volumeneinheit des Basisraumes gleich $e\, p/\tau_p$. Wird für p der Mittelwert der linearen Verteilung von Abb. 6.7 als $p(0)/2$ genommen und mit dem Volumen der Basiszone $F \cdot W$ multipliziert, so erhält man als Rekombinationsstrom

$$I_{pV} = \frac{e\, F\, W\, p_N}{2\tau_p}\, e^{U_{EB'}/U_T} \qquad \text{(Gl. 6.8c)}$$

Wird wieder der Elektronenanteil des Emitterstromes vernachlässigt, so erhalten wir wegen Gl. 6.1 mit Gl. 6.8a, 6.8c und 5.51 die Stromverstärkung (siehe auch Abb. 6.2).

$$A_N = -\frac{I_C - I_{CBO}}{I_E} \approx \frac{I_{Ep} - I_{pV}}{I_{Ep}} = 1 - \frac{W^2}{2L_p^2} \qquad \text{(Gl. 6.8d)}$$

Für einen typischen Niederfrequenz-Transistor ist $W \approx 40\,\mu$, $\tau_p \approx 15\,\mu\text{sec}$. Dies ergibt mit dem Wert $D_p = 48\ \text{cm}^2/\text{sec}$ aus Tab. 4 eine Stromverstärkung $A_N = 0{,}99$, also nahezu gleich 1.

Für eine genauere Analyse sind die Elektronenströme und die Rekombination von vornherein zu berücksichtigen. Wir müssen deshalb die Differentialgleichung 6.5 mit den Randbedingungen Gl. 6.6, 6.7 exakt lösen. Dies geschieht mit der allgemeinen Lösung Gl. 5.50 und ergibt für $0 < x < W$:

$$\delta p(x) = \frac{p_N}{\sinh(W/L_p)} \left\{ (e^{U_{EB'}/U_T} - 1) \sinh\left(\frac{W - x}{L_p}\right) - \sinh\left(\frac{x}{L_p}\right) \right\} \qquad \text{(Gl. 6.8)}$$

Für eine sehr enge Basis $W \ll L_p$, also für den oben betrachteten Näherungsfall ergibt das:

$$p(x) = p_N + \delta p = p_N\, e^{U_{EB'}/U_T} \left(1 - \frac{x}{W}\right) \qquad \text{(G. 6.9)}$$

Dies ist die bereits erwähnte und in Abb. 6.7 dargestellte lineare Verteilung.

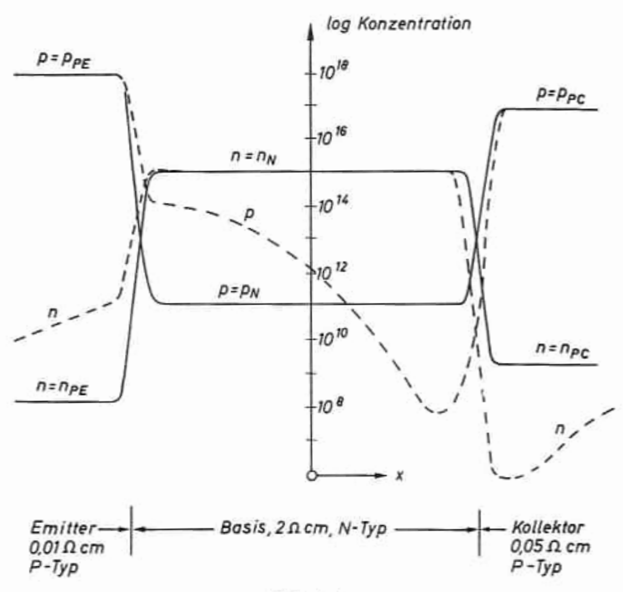


Abb. 6.8.
Die Trägerdichten im PNP-Transistor: ——— bei Gleichgewicht ($U_{EB'} = 0, U_{CB'} = 0$), -- im Verstärkerbetrieb ($U_{EB'} > 0, U_{CB'} < 0$).

In Abb. 6.8 ist der Verlauf der Trägerkonzentrationen noch einmal im logarithmischen Maßstab entsprechend Abb. 5.31 aufgetragen. Man sieht, wie die gestrichelte Verteilung der Löcherdichte durch Anheben am Emitter ($U_{EB'} > 0$) und Absenken am Kollektor ($U_{CB'} > 0$) zustande kommt. Die ausgezogenen Kurven gelten für $U_{EB'} = U_{CB'} = 0$.

Die Elektronendichte in der Basis ist wegen der Ladungsneutralität der Diffusionszone in der Basis für kleine Ströme (kleine δp)

$$0 < x < W: \qquad n(x) = n_N + \delta p(x) \approx n_N \tag{Gl. 6.10}$$

Die Berechnung der Elektronendichte in der Emitter- und Kollektorzone können wir vom ungestörten PN-Problem, Abschnitt 5.4.3.4, übernehmen. Die dortige Gl. 5.60 gibt, sinngemäß auf Elektronen übertragen, wenn L_{nE} und L_{nC} die Diffusionslängen der Elektronen in der Emitter- bzw. Kollektorzone sind:

$$x < 0: \qquad \delta n(x) = n_{PE}\left(e^{U_{EB'}/U_T} - 1\right) e^{x/L_{nE}} \tag{Gl. 6.11}$$

und

$$x > W: \qquad \delta n(x) = -\, n_{PC}\, e^{-(x-W)/L_{nC}} \tag{Gl. 6.12}$$

Aus den Minoritätsträgerdichten ergeben sich dann die Ströme als Diffusionsströme an den Feldzonenrändern. Es wird der Löcheranteil des Emitter-Stromes, wenn F die Fläche des PN-Übergangs ist:

$$I_{Ep} = -\, e F D_p \left.\frac{\partial \delta p}{\partial x}\right|_{x=0} \tag{Gl. 6.13}$$

$$= \frac{e F D_p p_N}{L_p \sinh(W/L_p)} \left\{ \left(e^{U_{EB'}/U_T} - 1 \right) \cosh\left(\frac{W}{L_p}\right) + 1 \right\}$$

Mit der Näherung von Gl. 6.9 erhält man natürlich den Ausdruck Gl. 6.8a.

Der Elektronenanteil des Emitterstromes ist

$$I_{En} = +\, e F D_n \left.\frac{\partial \delta n}{\partial x}\right|_{x=0} = \frac{e F D_n n_{PE}}{L_{nE}} \left(e^{U_{EB'}/U_T} - 1\right) \tag{Gl. 6.14}$$

Ferner sind die beiden Anteile des Kollektorstromes (positiv gezählt in Richtung $-x$!)

$$I_{Cp} = +\, e F D_p \left.\frac{\partial \delta p}{\partial x}\right|_{x=W}$$

$$= -\frac{e F D_p p_N}{L_p \sinh\left(\frac{W}{L_p}\right)} \left\{ \left(e^{U_{EB'}/U_T} - 1\right) + \cosh\left(\frac{W}{L_p}\right) \right\} \tag{Gl. 6.15}$$

und

$$I_{Cn} = -\ e F D_n \frac{\partial \delta n}{\partial x}\bigg|_{x=W} = -\ \frac{e F D_n n_{PC}}{L_{nC}} \tag{Gl. 6.16}$$

Und schließlich erhalten wir durch Zusammenfassen dieser Gleichungen die Stromspannungs-Beziehungen des inneren Transistors:

$$I_E = a_{11}\left(e^{U_{EB'}/U_T} - 1\right) - a_{12} \tag{Gl. 6.17}$$

$$I_C = a_{21}\left(e^{U_{EB'}/U_T} - 1\right) - a_{22} \tag{Gl. 6.18}$$

mit den Koeffizienten

$$a_{11} \equiv e F \left(\frac{D_p p_N}{L_p \tanh\left(\frac{W}{L_p}\right)} + \frac{D_n n_{PE}}{L_{nE}}\right) \tag{Gl. 6.19}$$

$$a_{12} = a_{21} \equiv -\ e F \frac{D_p p_N}{L_p \sinh\left(\frac{W}{L_p}\right)} \tag{Gl. 6.20}$$

$$a_{22} \equiv e F \left(\frac{D_p p_N}{L_p \tanh\left(\frac{W}{L_p}\right)} + \frac{D_n n_{PC}}{L_{nC}}\right) \tag{Gl. 6.21}$$

Ein Vergleich dieser Formeln mit Gl. 5.54 und 5.55 zeigt, daß die Größen a_{11}, a_{12}, a_{22} im wesentlichen die Diodensättigungsströme sind. Die Löcheranteile sind lediglich wegen der endlichen Dicke W der Basis durch die Größen $\sinh(W/L_p)$ bzw. $\tanh(W/L_p)$ modifiziert.

Eliminiert man aus Gl. 6.17 und 6.18 die Emitter-Spannung, so erhält man

$$I_C = I_{CBO} - A_N I_E \tag{Gl. 6.22}$$

was mit Gl. 6.1 identisch ist.

Hierbei ist jetzt:

$$I_{CBO} \equiv \frac{a_{12}\, a_{21} - a_{11}\, a_{22}}{a_{11}} \tag{Gl. 6.23}$$

und

$$A_N = -\frac{a_{12}}{a_{11}} \tag{Gl. 6.24}$$

Durch Einsetzen von Gl. 6.19 und 6.20 in Gl. 6.24 sieht man, daß sich die Stromverstärkung A_N als Produkt zweier Faktoren ausdrücken läßt

$$A_N = \frac{1}{\cosh\left(\frac{W}{L_p}\right)} \cdot \frac{\frac{D_p p_N}{L_p}}{\frac{D_p p_N}{L_p} + \frac{D_n n_{PE}}{L_{nE}} \tanh\left(\frac{W}{L_p}\right)} \equiv A_0 \cdot \gamma \quad \text{(Gl. 6.25)}$$

Der Faktor A_0 ist der „Transportfaktor" oder „Rekombinationsfaktor". Er beschreibt die Auswirkung des Rekombinationsverlustes in der Basis auf die Stromverstärkung. Für einen guten Transistor ist immer $W \ll L_p$, so daß entsprechend Gl. 6.8d gilt:

$$A_0 \approx 1 - \frac{W^2}{2\,L_p^2} \quad \text{(Gl. 6.26)}$$

Der zweite Faktor γ kann mit Hilfe von Gl. 6.13 und 6.14 als Anteil des Löcherstromes am gesamten Emitterstrom aufgefaßt werden, wenn in Gl. 6.13 das Glied $+1$ in der Klammer (also der Strom bei $U_{EB'} = 0$ gegenüber dem Gesamtstrom) vernachlässigt wird.

$$\gamma = \frac{I_{Ep}}{I_{Ep} + I_{En}} = \frac{1}{1 + I_{En}/I_{Ep}} \quad \text{(Gl. 6.27)}$$

Man nennt diese Größe deshalb „Emitterwirkungsgrad". Sie kann durch hohe P-Dotierung des Emitters ($p_{PE} \gg n_N$, also $n_{PE} \ll p_N$) ebenfalls nahezu Eins gemacht werden. Dies gilt jedoch nur für kleine Ströme, worauf wir in Abschnitt 6.6 noch zu sprechen kommen werden.

Für kleine Ströme ist I_{En}/I_{Ep} von der Größenordnung n_{PE}/p_N, da die anderen Faktoren dieses Bruches sich etwa wegheben. Für die in Abb. 6.8 angegebenen Dichten ist dies 10^{-3}, was $\gamma \approx 0{,}999$ ergibt. Man sieht also, daß die Näherung von Gl. 6.8a (d. h. $\gamma = 1$) hierfür berechtigt ist.

I_{CBO} ist der Kollektorsättigungsstrom bei offenem Emitter, auch Kollektorreststrom genannt. Durch Einsetzen von Gl. 6.19 bis 6.21 in Gl. 6.23 erhält man einen längeren Ausdruck, den wir vereinfachen können, wenn die oben erwähnten Voraussetzungen für einen guten Transistor, $W \ll L_p$ und $n_{PE} \ll p_N$, erfüllt sind. Man erhält dann (unter Berücksichtigung von Gl. 5.51)

$$I_{CBO} \approx -F\left(e\frac{p_N}{\tau_p}W + e\,D_n\frac{n_{PC}}{L_{nC}}\right) \quad \text{(Gl. 6.28)}$$

Das erste Glied dieses Ausdruckes stellt nach den Ausführungen im Anschluß an Gl. 5.54 und nach Gl. 5.55 den Strom dar, der durch Generation der Löcher in der Basis (Dicke W) entsteht, während das zweite Glied nach Gl. 6.16 den im Kollektorgebiet erzeugten Elektronenstrom darstellt (vgl. Abb. 6.2).

An dieser Stelle soll gleich noch die Stromverstärkung im inversen Betrieb definiert werden, d. h. bei Vertauschung von Emitter und Kollektor. Eine Betrachtung von Gl. 6.17 bis 6.21 und 6.24 zeigt, daß diese Vertauschung als inverse Stromverstärkung

$$A_I = -a_{12}/a_{22} \qquad \text{(Gl. 6.24a)}$$

liefert.

Aus Gl. 6.22 und Gl. 6.28 geht hervor, daß für Kollektorsperrspannungen $-U_{CB'} \gg U_T$, d. h. größer als etwa 100 mV, der Kollektorstrom von der Kollektorspannung unabhängig ist, d. h. die $I_C - U_{CB'}$-Kennlinien für konstantes I_B, wie sie in Abb. 6.3 dargestellt sind, sollten horizontal verlaufen. Mit anderen Worten, der Ausgangswiderstand r_C muß beim idealisierten Transistor unendlich groß sein.

Dies ist bei einem wirklichen Transistor jedoch nicht der Fall; r_C hat eine endliche Größe. Um dies zu erklären, muß das Modell des idealisierten inneren Transistors ergänzt werden. Dies soll im folgenden Abschnitt geschehen.

6.3.2 Basisdickenänderung, Early-Effekt. Wir haben im vorangegangenen Abschnitt stillschweigend so getan, als ob die Basisdicke W eine Konstante wäre. Tatsächlich ist die Basis aber durch die beiden Sperrschichten begrenzt, deren Dicke nach Gl. 5.30 eine Funktion der anliegenden Spannungen ist. Nun ist bei normalem Betrieb allerdings die Emitterspannung klein, so daß sich nach Gl. 5.30 eine Änderung derselben nicht sehr stark auswirkt. An der Kollektorsperrschicht liegt jedoch eine große Spannung, und wir haben deren Einfluß auf die Dicke der Kollektorsperrschicht l_C und damit auf die Basisdicke W zu betrachten. Dies ergibt eine Abhängigkeit der Ströme I_E und I_C von $U_{CB'}$, ein Effekt, der zuerst von Early [78] beschrieben wurde.

Wir haben also:

$$W(U_{CB'}) = W_0 - l_C(U_{CB'}) = W_0 - l_0 \sqrt{1 - \frac{U_{CB'}}{V_D}}$$

$$\approx W_0 - l_0 \sqrt{\frac{-U_{CB'}}{V_D}} \qquad \text{(Gl. 6.29)}$$

Die letzten beiden Gleichungsteile beziehen sich auf den Fall eines legierten Kollektors, bei dem der Übergang abrupt und $n_N \ll p_{PC}$ ist. W_0 ist die Dicke der gesamten N-Zone. Die Dicke der Emittersperrschicht ist klein und kann vernachlässigt werden. So erhält man als Änderung der Basisdicke W mit der Kollektorspannung $-U_{CB'}$

$$\frac{\partial W}{\partial U_{CB'}} = \frac{l_0}{2\sqrt{V_D \cdot (-U_{CB'})}} \qquad \text{(Gl. 6.30)}$$

Dies ergibt mit den Werten von Abschnitt 5.4.2.1 und Abb. 5.22 ($l_0 = 0{,}76\,\mu$, $V_D = 0{,}27$ V) bei $-U_{CB'} = 5$ V eine Änderung von $0{,}33\,\mu/\text{V}$, einen Wert also, der durchaus ins Gewicht fallen kann.

Für andere Strukturen, wie z. B. den Mesatransistor, ist mit Hilfe der vollständigen Gl. 5.30 bzw. 5.21a sinngemäß zu verfahren.

Gl. 6.29 ergibt dann, in die Gl. 6.17 bis 6.21 eingesetzt, die verbesserten Stromspannungsgleichungen. Diese werden kompliziert und deshalb hier nicht explizite angeschrieben. Wir wollen lediglich die Neigung der Kollektorkennlinien $1/r_C$ (Abb. 6.3), d. h. den Ausgangswiderstand in einem vorgegebenen Gleichstromarbeitspunkt I_C und bei festgehaltener Emitterspannung abschätzen. Es soll wieder $W \ll L_p$ und $n_N \ll p_{PC}$ sein, so daß wir von Gl. 6.8a ausgehen können. Es ist dann

$$1/r_C = \left(\frac{\partial I_C}{\partial U_{CB'}}\right)_{U_{EB'}} = -\left(\frac{\partial I_E}{\partial U_{CB'}}\right)_{U_{EB'}} \approx + \frac{I_E}{W} \frac{\partial W}{\partial U_{CB'}} \quad \text{(Gl. 6.31)}$$

Für $I_C \approx I_E = 1$ mA und $W = 40\,\mu$ erhält man mit der obigen Zahl für $(\partial W/\partial U_{CB'})$ einen Ausgangswiderstand $r_C \approx 120\,\mathrm{k}\Omega$, ein Wert, der in Abschnitt 6.1 schon benützt wurde.

6.3.3 Basiswiderstand, Gleichstrom-Ersatzschaltbild. In den Gleichungen der beiden vorhergegangenen Abschnitte sind stets nur die Spannungen an den beiden PN-Übergängen aufgetreten, die wir $U_{EB'}$ bzw. $U_{CB'}$ genannt haben. Mit anderen Worten: wir haben uns bisher nur mit dem „inneren“ Transistor befaßt. Tatsächlich sind zu diesen Spannungen noch die Spannungsabfälle an den Bahnwiderständen in den Emitter-, Kollektor- und Basiszonen hinzuzuzählen, um den zu äußeren Spannungen zwischen Emitter- und Basisklemme

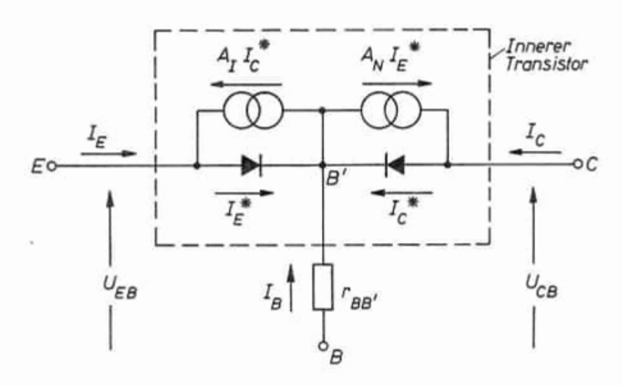


Abb. 6.9.
Das Gleichstrom-Ersatzschaltbild des Transistors.

U_{EB} bzw. zwischen Kollektor- und Basisklemme U_{CB} („B“ ohne „Strich“!) zu kommen. Man kann allerdings bei den meisten Transistortypen, insbesondere bei Legierungstransistoren, wo p_{PE} und $p_{PC} \gg n_N$ ist, die Emitter- und Kollektor-Bahnwiderstände vernachlässigen. Nicht zu vernachlässigen ist aber der

Basis-Bahnwiderstand, da es sich hier um einen Querwiderstand in einer dünnen Schicht handelt.

Wir können also den realen Transistor annähern durch das Schaltbild Abb. 6.9. Der Schaltungsteil innerhalb des gestrichelt gezeichneten Kastens stellt den in Abschnitt 6.3.1 und 6.3.2 behandelten inneren Transistor dar. Dieser ist nun ergänzt durch den Basiswiderstand $r_{BB'}$.

Man hat demnach

$$U_{EB} = U_{EB'} - r_{BB'} I_B \quad \text{(Gl. 6.34)}$$

und

$$U_{CB} = U_{CB'} - r_{BB'} I_B \quad \text{(Gl. 6.35)}$$

Diese beiden Gleichungen stellen, zusammen mit Gl. 6.17 und 6.18 (mit Gl. 6.19 bis 6.21 und 6.29), eine vollständige Beschreibung der Gleichstromkennlinien des „idealisierten, äußeren" Transistors dar. Sie lassen sich durch das Ersatzschaltbild Abb. 6.9 wiedergeben, wenn wir die Ströme I_E^* und I_C^* durch die „Emitter"- und „Kollektor-Diode" definieren als

$$I_E^* = a_{11} \left(e^{U_{EB'} / U_T} - 1\right) \quad \text{(Gl. 6.36)}$$

und

$$I_C^* = - a_{22} \quad \text{(Gl. 6.37)}$$

Die Äquivalenz des Ersatzschaltbildes mit den oben erwähnten Kennliniengleichungen ist unter Benützung von Gl. 6.24, 6.24a, 6.36 und 6.37 leicht zu zeigen.

Wir sehen, daß die Wechselwirkung zwischen den Strömen durch die beiden *PN*-Dioden, die ja bekanntlich den Verstärkereffekt zur Folge hat, durch die beiden Stromgeneratoren $A_N I_E^*$ und $A_I I_C^*$ beschrieben werden kann. Diese Stromgeneratoren machen den Transistorvierpol zum aktiven Vierpol.

Als nächstes wollen wir den Basiswiderstand $r_{BB'}$ berechnen. Diese Aufgabe wird kompliziert durch die Tatsache, daß die verschiedenen Teile des Basisstromes auf verschiedenen Strompfaden laufen und deshalb ganz verschiedene Spannungsabfälle erzeugen. Wir haben also einen verteilten Widerstand. Ferner sind die Stromverteilungen der verschiedenen Komponenten des Basisstromes, wie z. B. $(1 - A_0) I_{Ep}$, I_{En} und I_{CBO} (siehe Abb. 6.2), verschieden. Dies ist am Beispiel eines Legierungstransistors in Abb. 6.10 dargestellt. Dabei stellt der Teil *a* die Stromverteilung des Rekombinationsstromes $(1 - A_0) I_{Ep}$ dar, während in Teil *b* und *c* der Elektronenanteil des Emitterstromes I_{En} bzw. der Kollektorreststrom I_{CBO} eingetragen sind. Letzterer ist jedoch in der Regel klein gegenüber den beiden anderen Komponenten. Zudem können die Fälle *a* und *b* als gleich betrachtet werden, wenn die Basisdicke W klein ist gegenüber den anderen Dimensionen, z. B. gegenüber dem Emitterradius a, dem Kollektorradius b und dem inneren Radius des Basisringes c. Die relative Bedeutung der verschiedenen Komponenten und die Stromverteilungen sind

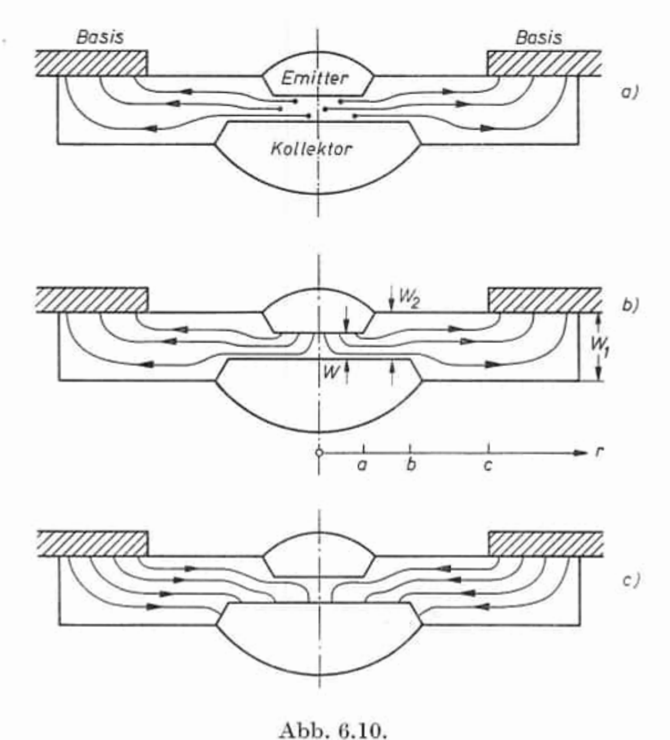


Abb. 6.10.
Stromverteilungen des Basisstromes: a) Rekombinationsstrom $(1 - A_D)\, I_{Ep}$, b) Elektronenstrom, I_{En}, c) Kollektorreststrom, I_{CB0}.

natürlich für jeden Transistortyp verschieden. Wir wollen im folgenden lediglich den in Abb. 6.10 gezeichneten Legierungstransistor als Beispiel durchrechnen. Für andere Typen muß sinngemäß entsprechend verfahren werden.

Der Basiswiderstand $r_{BB'}$ ist definiert durch die Gl. 6.34 als der Spannungsabfall, um den die Klemmenspannung U_{EB} verringert werden muß, um die Spannung am Emitterübergang $U_{EB'}$ zu erhalten, geteilt durch den Basisstrom. Aus Abb. 6.10 geht unmittelbar hervor, daß verschiedene Teile der Emitterfläche verschiedene Spannungsabfälle sehen werden. Wir haben also einen geeigneten Mittelwert $\overline{U}$ zu finden.

An einer beliebigen Stelle der Emitterfläche F ist die Emitterstromdichte

$$j_E \approx \text{const} \cdot e^{\frac{U_{EB'}}{U_T}} = \text{const} \cdot e^{\frac{U_{EB} - U(r)}{U_T}}$$

wobei $U(r)$ der vom Ort abhängige Spannungsabfall am Basiswiderstand ist.

Für kleine Ströme ist $\overline{U}$ und $U(r) \ll U_T$ und damit

$$I_E = \int_0^F j_E\, dF = \text{const} \cdot e^{\frac{U_{EB}}{U_T}}\, F\left(1 - \frac{1}{F U_T}\int_0^F U(r)\, dF\right) \qquad \text{(Gl. 6.38)}$$

Andererseits ist der Mittelwert gegeben durch

$$I_E = \text{const} \cdot F\, e^{\frac{U_{EB}-\overline{U}}{U_T}} = \text{const} \cdot F\, e^{\frac{U_{EB}}{U_T}}\left(1 - \frac{\overline{U}}{U_T}\right) \qquad \text{(Gl. 6.39)}$$

Durch Vergleich von Gl. 6.38 mit 6.39 ergibt sich

$$r_{BB'} = \frac{\overline{U}}{I_B} = \frac{1}{F I_B}\int_0^F U(r)\, dF = \frac{2}{a^2 I_B}\int_0^a U(r)\, r\, dr \qquad \text{(Gl. 6.40)}$$

Zur Berechnung der Spannung $U(r)$ dient das Ohmsche Gesetz für zylindrischen Stromfluß $\left(\boldsymbol{F} = -\operatorname{grad} U = \frac{\boldsymbol{j}}{\sigma}\right)$

$$\frac{d\,U(r)}{d\,r} = -\frac{I_B(r)}{2\pi r W(r)} \cdot \frac{1}{\sigma_N} \qquad \text{(Gl. 6.41)}$$

Wir haben dabei drei Bereiche, die zu dieser Spannung beitragen (siehe Abb. 6.10b):

1. der Ring $c > r > b$ mit $W(r) = W_1$ und $I_B(r) = I_B$,
2. der Ring $b > r > a$ mit $W(r) = W_2$ und $I_B(r) = I_B$,
3. der Kreiszylinder $a > r > 0$ mit $W(r) = W$ und $I_B(r)$ proportional zum Volumen, von $r = 0$ aus wachsend gemäß

$$dI_B(r) = \frac{I_B}{W\pi a^2} W\, 2\pi r\, dr$$

und damit

$$I_B(r) = I_B \frac{r^2}{a^2} \qquad \text{(Gl. 6.42)}$$

Die Integration von Gl. 6.41 von $r = c$ durch den ersten und zweiten Bereich hindurch bis zu einem Punkt im dritten Bereich, also von $U(r{=}c) = 0$ bis $U(r < a)$ ergibt

$$U(r) = \frac{I_B}{2\pi\sigma_N}\left(\frac{\ln c/b}{W_1} + \frac{\ln b/a}{W_2} + \frac{1 - r^2/a^2}{2W}\right)$$

Durch Einsetzen in Gl. 6.40 und Integration über die Emitterfläche, d. h. von $r = 0$ bis $r = a$ erhält man schließlich

$$r_{BB'} = \frac{1}{2\pi\sigma_N}\left(\frac{\ln c/b}{W_1} + \frac{\ln b/a}{W_2} + \frac{1}{4W}\right) \qquad \text{(Gl. 6.43)}$$

Als Beispiel wollen wir die Zahlenwerte der Type OC 304 verwenden mit $a = 425\,\mu$, $b = 500\,\mu$, $c = 660\,\mu$, $W_1 = 90\,\mu$, $W_2 = 65\,\mu$, $W = 40\,\mu$ und $\sigma_N = 0{,}2\,\Omega^{-1}\text{cm}^{-1}$. Man erhält für die drei Anteile des Basiswiderstandes aus Gl. 6.43 $r_{BB'} = 25 + 20 + 115\,\Omega = 160\,\Omega$.

Dieser Wert kommt schon in eine Größenordnung, wo er den Eingangswiderstand des Transistors vergrößern und damit die Verstärkung herabsetzen kann. Er stellt in der Schaltung Abb. 6.1 eine Gegenkopplung zwischen Ausgang und Eingang dar. Insbesondere aber trägt ein großes $r_{BB'}$ wesentlich zur Verschlechterung des Hochfrequenzverhaltens bei. Aus diesen Gründen versucht man, den Basiswiderstand so klein wie möglich zu machen. Dies kann, wie Gl. 6.43 zeigt, z. B. durch Verkleinerung des Abstandes zwischen Emitterumfang und Basiskontakt $(c - a)$ geschehen, was aus fertigungstechnischen Gründen eine Grenze hat. Diese Grenze ist bei dem angeführten Beispiel schon erreicht. Überdies wird auch $r_{BB'}$ hier schon im wesentlichen durch das Mittelglied bestimmt, das nur durch eine Vergrößerung von W verkleinert werden kann. Damit wird jedoch die Stromverstärkung A_N und insbesondere die Frequenzgrenze wieder schlechter. Schließlich kann die Dotierung in der Basis vergrößert werden, wodurch sich aber jetzt der Emitterwirkungsgrad γ verschlechtert. Man sieht an diesem Beispiel sehr deutlich, wie der Bau eines für einen bestimmten Zweck geeigneten Transistors die sehr sorgfältige Wahl und Erprobung von Kompromissen nötig macht.

Für die Frequenzgrenze von HF-Transistoren ist auch der Verschiebungsstrom über die Kollektorkapazität maßgebend. Dieser ist aber dann nicht durch die Stromverteilung von Abb. 6.10a oder b, sondern durch diejenige von Abb. 6.10c gegeben. Dementsprechend sieht nur der vom Mittelteil des Kollektors kommende Strom den oben beschriebenen Basiswiderstand, während der äußere Teil des Kollektors über einen sehr viel kleineren Widerstand mit dem Basiskontakt verbunden ist (vgl. Abschnitt 6.4.3).

6.3.4 Das dreidimensionale Transistorproblem. Wir haben zwar im letzten Abschnitt den Basiswiderstand auf Grund eines realistischen dreidimensionalen Modells ausgerechnet, aber unser „innerer“ Transistor ist bisher nur als eindimensionales Problem behandelt worden. Um zum realen Transistor zu kommen, haben wir die Oberflächenrekombination an der Oberfläche der Basiszone zu berücksichtigen. Außerdem ist je nach dem Transistortyp die Abweichung der Geometrie vom eindimensionalen Fall zu berücksichtigen. In dieser Hinsicht wäre der „gezogene“ Transistor (Abb. 5.14) unserer bisherigen Rechnung

am ähnlichsten. Wir wollen uns aber wie im letzten Abschnitt auf das Beispiel des legierten Transistors beschränken (Abb. 5.17 und 6.10), da ihm größere Bedeutung zukommt.

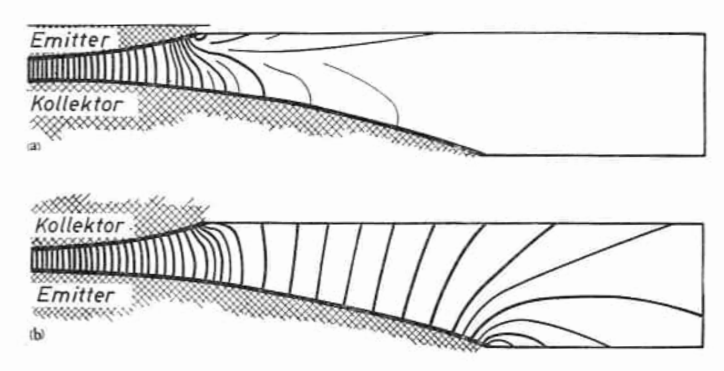


Abb. 6.11.
Stromverteilung des Minoritätenstromes im dreidimensionalen Modell des Legierungstransistors a) Normaler Betrieb (Emitter kleiner als Kollektor), b) Inverser Betrieb (Emitter größer als Kollektor). Nach MOORE und PANKOVE [79].

In Abb. 6.11 sind die Stromlinien des Minoritätenstromes für einen solchen Fall aufgezeichnet [79]. Und zwar ist in a) der Emitter kleiner als der Kollektor. In b) ist es umgekehrt. Man sieht, daß ein Teil der Löcher zur Oberfläche geht und dort rekombiniert, und daß dieser Effekt in a) weit geringer ist als in b).

Die mathematische Behandlung dieses Problems [80, 81] läuft auf eine Lösung der dreidimensionalen Bilanzgleichung

$$\Delta\, \delta p = \frac{\delta p}{L_p^2} \qquad \text{(Gl. 6.44)}$$

in zylindrischen Koordinaten hinaus.

Dabei sind als Randbedingungen an der Emitterfläche Gl. 6.6, an der Kollektorfläche Gl. 6.7 bzw. $p \approx 0$ und an den freien Oberflächen nach Gl. 4.82

$$\text{grad}_\perp\, \delta p = -\frac{s}{D_p}\, \delta_p \qquad \text{(Gl. 6.45)}$$

zu nehmen. $\text{grad}_\perp$ ist der Gradient in Richtung senkrecht auf die Oberfläche zu.

Die Rechnung wird recht kompliziert, und das Ergebnis ist schlecht überschaubar. Eine Vereinfachung wird erreicht durch Vernachlässigung der Volumenrekombination gegenüber der Oberflächenrekombination, was bei den meisten Transistoren gerechtfertigt ist. Dann wird aus Gl. 6.44 die LAPLACE-Gleichung

$$\Delta\, \delta p = 0 \qquad \text{(Gl. 6.44a)}$$

Die Stromlinien können nun mit Hilfe des Strom-Analog-Verfahrens im elektrolytischen Trog [82] bzw. für zweidimensionale Probleme auf dem Widerstandspapier [79] ermittelt werden. Nach dem letzteren Verfahren ist auch die Lösung in Abb. 6.11 entstanden. Das Ergebnis dieser Untersuchungen ist unter anderem, daß die Stromverstärkung am größten wird, wenn der Kollektordurchmesser etwas mehr als doppelt so groß wie der Emitterdurchmesser gemacht wird.

Für eine qualitative Übersicht eignet sich besser eine Näherung, die in Abb. 6.12 angedeutet ist [83]. Es wird dabei für den Zylinder unterhalb des Emitters der Fläche πa^2 ein Stromfluß mit entsprechender Rekombination angenommen (eindimensionale Behandlung). Für die Oberflächenrekombination wird vorausgesetzt, daß sie in einer Ringfläche der Breite b^* erfolgt, die den Emitter umschließt und an der die Löcherdichte gleich derjenigen direkt am Emitterübergang sein soll.

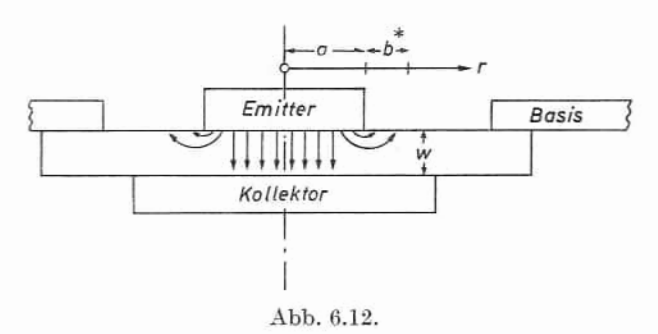


Abb. 6.12.

Zur näherungsweisen Berechnung des dreidimensionalen Transistors.

Für nicht zu kleinen Emitterstrom und gesperrten Kollektor kann in Gl. 6.22 das erste Glied der rechten Seite (I_{CBO}) vernachlässigt werden, und man hat

$$A_N = -\frac{I_C}{I_E} \tag{Gl. 6.46}$$

Außerdem ist (vgl. Abb. 6.2)

$$-I_C = I_E - I_{En} - I_{pV} - I_{pO}$$

$$= I_E - e\pi a^2 \frac{D_n}{L_{nE}} n_{PE}\, e^{\frac{U_{EB'}}{U_T}} - e\pi a^2 \frac{W}{2\,\tau_p} p_N\, e^{\frac{U_{EB'}}{U_T}}$$

$$- e\, 2\pi a b^* s\, p_N\, e^{\frac{U_{EB'}}{U_T}} \tag{Gl. 6.47}$$

Hier ist I_{pV} der Stromverlust durch Volumenrekombination; er wird auf Grund der linearen Löcherverteilung Gl. 6.9 und Abb. 6.7 entsprechend Gl. 5.55 berechnet. I_{pO} ist der Stromverlust durch Oberflächenrekombination und wird mit $\delta p \approx p_N \, e^{U_{EB'}/U_T}$ nach Gl. 4.82 berechnet. Wenn dies in Gl. 6.46 eingesetzt und für I_E der Ausdruck von Gl. 6.8a verwendet wird, erhält man

$$A_N = 1 - \frac{\sigma_N W}{\sigma_{PE} L_{nE}} - \frac{W^2}{2 L_p^2} - \frac{2 b^* W s}{a D_p} \qquad \text{(Gl. 6.48)}$$

Das erste und dritte Glied sind die wohlbekannte Gl. 6.26 des eindimensionalen Modells. Der zweite Term ist der Einfluß der Emitterergiebigkeit, und der vierte Term beschreibt die Oberflächenrekombination. Man sieht, daß A_N umso kleiner wird, je größer die Oberflächenrekombinations-Geschwindigkeit s, je kleiner der Emitterradius a und je größer die Basisdicke W wird. Letzteres gilt auch noch, wenn W klein genug ist, um den Volumenrekombinationsterm klein gegen Eins zu machen.

Die beiden Rekombinationsterme können auch mit Hilfe einer effektiven Lebensdauer τ_{eff} zusammengefaßt werden:

$$\frac{W^2}{2 D_p \tau_p} + \frac{2 b^* W s}{a D_p} \equiv \frac{W^2}{2 D_p \tau_{eff}}$$

Daraus ergibt sich eine effektive Lebensdauer zu

$$\frac{1}{\tau_{eff}} = \frac{1}{\tau_p} + \frac{4 b^*}{a W} s \qquad \text{(Gl. 6.49)}$$

Wenn man gemäß Abb. 6.11a die Breite b^* zu $W/2$ ansetzt, so zeigt der Vergleich mit der oberen Hälfte von Gl. 4.86, daß der dortigen Plattendicke (dort $2a$ genannt) hier der Emitterradius a entspricht. Man kommt also zu dem Schluß, daß in erster Näherung der dreidimensionale Transistor auf das eindimensionale Modell zurückgeführt werden kann, wenn statt der Volumenlebensdauer τ_p eine effektive Lebensdauer τ_{eff} von geeigneter Größe in die Gleichungen des eindimensionalen Modells eingesetzt wird.

6.4 Das Wechselstromverhalten bei kleinen Amplituden

Im vorliegenden Kapitel sollen die bisherigen Betrachtungen auf Wechselströme ausgedehnt werden. Dies wird zunächst in Abschnitt 6.4.1 für den „idealisierten" Transistor entsprechend Abschnitt 6.3.1 durchgeführt. In Abschnitt 6.4.2 wird diese Behandlung vervollständigt durch Berücksichtigung der Basisdickenänderung mit der Kollektorspannung. Wegen der Linearität

der Differentialgleichung und wegen der Kleinheit dieses Effektes kann dies lediglich durch Hinzufügung von weiteren Gliedern zu den Stromgleichungen geschehen. Man gelangt damit zu den Vierpolgleichungen des „inneren Transistors" und zum entsprechenden Ersatzschaltbild. In Abschnitt 6.4.3 soll dieses Ersatzschaltbild durch die Berücksichtigung des Basiswiderstandes und der Feldzonenkapazität beider *PN*-Übergänge ergänzt und entsprechende Vierpolkoeffizienten für den gesamten Transistor gefunden werden. Mit Hilfe dieser Koeffizienten lassen sich in Abschnitt 6.4.4 die Strom-, Spannungs- und Leistungsverstärkungen herleiten, und zwar zunächst für die basisgeerdete Schaltung, was dann in Abschnitt 6.4.5 auf die Emitterschaltung ausgedehnt wird.

6.4.1 Der idealisierte Transistor bei Wechselstrom. Wir verfahren entsprechend der Behandlung des einfachen *PN*-Übergangs in Abschnitt 5.6.2 und setzen die Spannungen an als Summe einer Gleich- und einer Wechselspannungskomponente (vgl. Gl. 5.78). Es sei also die Emitterspannung

$$(U_{EB'})_{\text{gesamt}} = U_{EB'} + u_{EB'}\, e^{i\omega t} \qquad \text{(Gl. 6.50)}$$

Um die Rechnung nicht unnötig zu komplizieren, soll wieder $u_{EB'} \ll U_T$ sein, was keine wesentliche Einschränkung für den normalen Verstärkerbetrieb bedeutet, da die Emitterspannung meist klein ist. Für die Kollektorspannung fordern wir lediglich, daß $-(U_{CB'})_{\text{ges.}} \gg U_T$, d. h. daß der Kollektor immer gesperrt ist. Wir erlauben also noch eine Kollektor-Wechselspannungsamplitude, die nur wenig kleiner als die Gleichspannung ist. Somit haben wir am Emitterübergang eine Löcherdichte gemäß Gl. 5.79.

$$p(x{=}0, t) = p_N\, e^{U_{EB'}/U_T} \left(1 + \frac{u_{EB'}}{U_T} e^{i\omega t}\right) \qquad \text{(Gl. 6.51)}$$

und am Kollektorübergang

$$p(x{=}W, t) = 0 \qquad \text{(Gl. 6.52)}$$

Diese Randbedingungen sind auf die Bilanzgleichung (mit dem Zeitdifferentialglied) anzuwenden. Wegen der Linearität dieser Gleichung kann wieder die Wechselstromkomponente $\tilde{p}$ der Löcherdichte getrennt berechnet und hniterher zur Gleichstromlösung Abschnitt 6.3.1 hinzugezählt werden. Es gilt also wie beim einfachen *PN*-Übergang (Gl. 5.80)

$$\frac{\partial \tilde{p}}{\partial t} = -\frac{\tilde{p}}{\tau_p} + D_p \frac{\partial^2 \tilde{p}}{\partial x^2} \qquad \text{(Gl. 6.53)}$$

mit

$$\tilde{p}(0, t) = p_N\, e^{U_{EB'}/U_T}\; \frac{u_{EB'}}{U_T} e^{i\omega t} \qquad \text{(Gl. 6.54)}$$

$$\tilde{p}(W, t) = 0$$

Als Lösung haben wir in Abschnitt 5.6.2 ebene Wellen der Form Gl. 5.81 kennen gelernt. Zur Befriedigung der neuen Randbedingungen benötigen wir beide Vorzeichen der Ausbreitungsgeschwindigkeit v. Wir setzen also an:

$$\tilde{p} = A\, e^{i\omega\left(t-\frac{x}{v}\right)} + B\, e^{i\omega\left(t+\frac{x}{v}\right)} = \left(A\, e^{\eta\frac{x}{L_p}} + B\, e^{-\eta\frac{x}{L_p}}\right) e^{i\omega t} \qquad \text{(Gl. 6.55)}$$

wo nach Gl. 5.83 (durch Einsetzen des Ansatzes in Gl. 6.53) die Größe η bestimmt wird.

$$\eta \equiv \sqrt{1 + i\,\omega\,\tau_p} \qquad \text{(Gl. 6.56)}$$

Die Anwendung der Randbedingungen Gl. 6.54 liefert

$$\tilde{p}(x,t) = \frac{p_N}{U_T}\, e^{U_{EB'}/U_T}\, \frac{\sinh\left[\eta\,(W-x)/L_p\right]}{\sinh\left[\eta\, W/L_p\right]}\, u_{EB'}\, e^{i\omega t} \qquad \text{(Gl. 6.57)}$$

Daraus ergibt sich analog zu Gl. 6.13 bzw. Gl. 6.15 der Löcheranteil des Emitterwechselstroms i_{Ep} und des Kollektorwechselstroms i_{Cp}. Die Elektronenanteile lassen sich aus der für Elektronen umgeschriebenen Gl. 5.97 für den einfachen PN-Übergang berechnen, wobei der Kollektorelektronenstrom $i_{Cn} = 0$ wird wegen $e^{-U_{CB'}/U_T} = 0$.

Man erhält schließlich für die Gesamtwechselstromamplituden (ohne den Faktor $e^{i\omega t}$):

$$i_E = i_{Ep} + i_{En}$$

$$= e\,F\left(\frac{D_p\, p_N}{L_p \tanh\;(\eta W/L_p)} + \frac{D_n\, n_{PE}}{L_{nE}}\right)\frac{\eta}{U_T}\, e^{U_{EB'}/U_T}\, u_{EB'}$$

$$= \tilde{a}_{11}\,\frac{\eta}{U_T}\, e^{U_{EB'}/U_T}\, u_{EB'} \qquad \text{(Gl. 6.58)}$$

$$i_C = -\,e\,F\left(\frac{D_p\, p_N}{L_p \sinh\;(W\eta/L_p)}\right)\frac{\eta}{U_T}\, e^{U_{EB'}/U_T}\, u_{EB'}$$

$$= \tilde{a}_{21}\,\frac{\eta}{U_T}\, e^{U_{EB'}/U_T}\, u_{EB'} \qquad \text{(Gl. 6.59)}$$

Hier sind $\tilde{a}_{11}$ und $\tilde{a}_{21}$ durch Gl. 6.19 bzw. Gl. 6.20 definiert, wenn in tanh und sinh die Argumente W/L_p durch $\eta\, W/L_p$ ersetzt werden.

Die in Gleichung 6.56 definierte Größe η ist komplex und hängt von der Frequenz ω ab. Dementsprechend werden die Koeffizienten der Gl. 6.58 und 6.59 komplex und frequenzabhängig sein.

In erster Näherung läßt sich leicht anschaulich zeigen, wie diese Frequenzabhängigkeit vom Emitter- und Kollektorstrom aussieht. Zu dem nieder-

frequenten Strom

$$i_E \approx i_C = \frac{\partial I_E}{\partial U_{EB'}} u_{EB'} \approx \frac{I_E}{U_T} u_{EB'}$$

nach Gl. 6.8b kommt noch ein Glied

$$+ i\omega C_{DE} u_{EB'}$$

als Leitwert der Diffusionskapazität hinzu. Diese ergibt sich ganz entsprechend Abschnitt 5.6.2.1 aus dem periodischen Aufbau und Abbau der Löcherdichte in der Basiszone. In Abb. 6.7 ist eine solche Änderung dp gestrichelt angedeutet. Es ist dann

$$C_{DE} = \frac{dQ_{\text{Basis}}}{dU_{EB'}} = \frac{d}{dU_{EB'}}\left(eFW\frac{p(0)}{2} \right)$$

und wegen Gl. 6.6 und dem Ausdruck für den Emitterstrom I_E Gl. 6.8a

$$C_{DE} = \frac{eFW\,p(0)}{2U_T} = \frac{W^2}{2D_p}\frac{I_E}{U_T} \qquad \text{(Gl. 6.60)}$$

$$i_E \approx i_C \approx \frac{I_E}{U_T}\left(1 + i\omega\frac{W^2}{2D_p}\right) u_{EB'} = \frac{I_E}{U_T}(1 + i\,f/f_L)\,u_{EB'} \qquad \text{(Gl. 6.61)}$$

Wie man sieht, wird die Diffusionskapazität entscheidend, wenn die Frequenz f in die Größenordnung der Frequenz $f_L = D_p/\pi W^2$ kommt. Wir nennen diese charakteristische Frequenz f_L Laufzeitgrenzfrequenz (vgl. Gl. 6.90). Sie liegt z. B. für einen typischen Niederfrequenztransistor mit $W = 40\,\mu$ (z. B. OC 304) und $D_p = 48$ cm²/sec bei $f = 1$ MHz. Gl. 6.61 kann durch Entwickeln des sinh in Gl. 6.59 für kleine Werte von $\eta W/L_p$ verifiziert werden.

6.4.2 Der Einfluß der Basisdickenänderung, die Leitwertsmatrix des „inneren" Transistors. Aus Gl. 6.58 und 6.59 geht hervor, daß beim idealisierten Transistor bei gesperrtem Kollektor, d. h. im normalen Verstärkerbereich, die Ströme nicht von der Wechselspannung $u_{CB'}$ am Kollektor abhängen. Tatsächlich zeigt sich beim realen Transistor ein zwar kleiner, aber endlicher Ausgangsleitwert. Die Ursache liegt wieder wie beim Gleichstrom in der in Abschnitt 6.3.2 beschriebenen Basisdickenänderung.

Um diesen Einfluß zu berücksichtigen, muß Gl. 6.53 noch einmal gelöst werden mit einer anderen Randbedingung bei $x = W$. Dies geschieht am einfachsten dadurch, daß, anstatt W zu ändern, die Löcherdichte p bei $x = W$ geändert wird. Dies ist in Abb. 6.13 dargestellt, wo der Löcherdichteverlauf

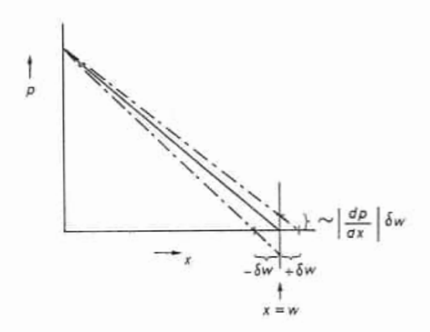


Abb. 6.13.
Minoritätendichte in der Basis bei Basisdickenänderung.

von Abb. 6.7 noch einmal als ausgezogene Linie gezeichnet ist. Die strichpunktierten Linien sind die Löcherdichte für $W + \delta W$ bzw. $W - \delta W$. Es ist aus der Figur leicht abzulesen, daß dann bei $x = W$ als Randbedingung

$$\tilde{p}(W, t) = -\left(\frac{\partial p}{\partial x}\right)_W \frac{\partial W}{\partial U_{CB'}} u_{CB'} e^{i\omega t} \qquad \text{(Gl. 6.62)}$$

anzusetzen ist. Da der Wechselstromanteil klein sein soll, kann $(\partial p/\partial x)_W$ in erster Näherung aus der Gleichstromverteilung Gl. 6.8 genommen werden. Die Löcherdichte $\tilde{p}$ am Emitter ($x = 0$) ändert sich nicht. Wir können die neuen Randbedingungen auch als eine Überlagerung von Gl. 6.54 (d. h. den Wechselstromfall mit $W =$ const.) und

$$\partial \tilde{p}(0, t) = 0$$

$$\delta\tilde{p}(W, t) = -\left(\frac{\partial p}{\partial x}\right)_W \frac{\partial W}{\partial U_{CB'}} u_{CB'} e^{i\omega t} \qquad \text{(Gl. 6.63)}$$

auffassen. Wegen der Linearität der Differentialgleichung Gl. 6.53 muß $\delta\tilde{p}$ mit Gl. 6.63 für sich diese Differentialgleichung befriedigen. Wir haben damit ein Problem, das demjenigen von Abschnitt 6.4.1 bis auf die Vertauschung von x mit $W - x$ völlig gleicht. Die Lösung $\delta\tilde{p}$ entspricht mit dieser Vertauschung Gl. 6.57, und die Ströme δi_E und δi_C entsprechen den Ausdrücken i_C nach Gl. 6.59 und dem Löcheranteil von i_E aus Gl. 6.58, wenn darin die Terme

$$\frac{p_N}{U_T} e^{U_{EB'}/U_T} u_{EB'}$$

durch

$$-\left(\frac{\partial p}{\partial x}\right)_W \frac{\partial W}{\partial U_{CB'}} u_{CB'}$$

ersetzt werden. $(\partial p/\partial x)_W$ kann aus Gl. 6.15 gewonnen werden. Außerdem soll

die Bezeichnung a_{12} von Gl. 6.20 benützt werden. Man erhält dann für $W \ll L_p$ die gesamten Wechselströme

$$i_E = i_E \text{ (Abschnitt 6.4.1)} + \delta i_E$$

$$= \tilde{a}_{11} \frac{\eta}{U_T} e^{U_{EB'}/U_T} u_{EB'} + a_{12} \frac{\eta}{L_p \sinh\left(\frac{\eta W}{L_p}\right)} e^{U_{EB'}/U_T} \frac{\partial W}{\partial U_{CB'}} u_{CB'} \quad \text{(Gl. 6.64)}$$

$$\equiv y'_{11} u_{EB'} + y'_{12} u_{CB'}$$

und

$$i_C = i_C \text{ (Abschnitt 6.4.1)} + \delta i_C$$

$$= \tilde{a}_{21} \frac{\eta}{U_T} e^{U_{EB'}/U_T} u_{EB'} - a_{22} \frac{\eta}{L_p \tanh\left(\frac{\eta W}{L_p}\right)} e^{\frac{U_{EB'}}{U_T}} \frac{\partial W}{\partial U_{CB'}} u_{CB'} \quad \text{(Gl. 6.65)}$$

$$\equiv y'_{21} u_{EB'} + y'_{22} u_{CB'}$$

Mit den letzten Teilen dieser Gleichungen haben wir die vier Koeffizienten der Leitwertsmatrix des inneren Transistors y'_{11}, y'_{12}, y'_{21} und y'_{22} definiert. Betrachten wir nämlich den Transistor in der Schaltung der Abb. 6.1 (Basisgeerdete Schaltung) als Vierpol, was in Abb. 6.14 angedeutet ist, so können Gl. 6.64 und 6.65 als die Vierpolgleichungen des inneren Transistors aufgefaßt werden.

$$i_E = y'_{11} u_{EB'} + y'_{12} u_{CB'} \quad \text{(Gl. 6.66)}$$

$$i_C = y'_{21} u_{EB'} + y'_{22} u_{CB'}$$

oder in Matrizenschreibweise

$$\begin{pmatrix} i_E \\ i_C \end{pmatrix} = \begin{pmatrix} Y' \end{pmatrix} \cdot \begin{pmatrix} u_{EB'} \\ u_{CB'} \end{pmatrix} \quad \text{(Gl. 6.67)}$$

wo

$$\begin{pmatrix} Y' \end{pmatrix} \equiv \begin{pmatrix} y'_{11} & y'_{12} \\ y'_{21} & y'_{22} \end{pmatrix} \quad \text{(Gl. 6.68)}$$

die Leitwertsmatrix des Vierpols ist.

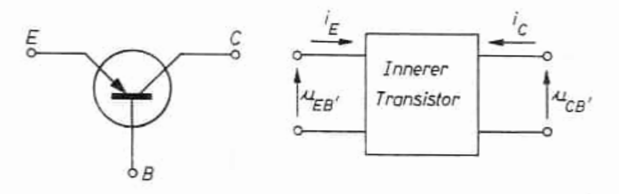


Abb. 6.14.
Der Transistor als Vierpol.

Um einen Eindruck von den Werten dieser Leitwertsparameter zu gewinnen, vernachlässigen wir — wie schon öfter — den Elektronenanteil des Emitterstroms und verwenden in den Größen a_{12}, $\tilde{a}_{11}$ und $\tilde{a}_{21}$ die Gl. 6.8a, so wird

$$\left(Y'\right) = I_E \begin{pmatrix} \frac{1}{U_T} \frac{\eta W/L_p}{\tanh(\eta W/L_p)} & -\frac{\eta W/L_p}{L_p \sinh(W/L_p) \sinh(\eta W/L_p)} \frac{\partial W}{\partial U_{CB'}} \\ -\frac{1}{U_T} \frac{\eta W/L_p}{\sinh(\eta W/L_p)} & \frac{\eta W/L_p}{L_p \sinh(W/L_p) \tanh(\eta W/L_p)} \frac{\partial W}{\partial U_{CB'}} \end{pmatrix}$$

(Gl. 6.68a)

Für $\eta W/L_p \ll 1$, also $W \ll L_p$ und tiefe Frequenz, vereinfacht sich dies weiter zu

$$\left(Y'\right) = I_E \begin{pmatrix} \frac{1}{U_T} & -\frac{1}{W} \frac{\partial W}{\partial U_{CB'}} \\ -\frac{1}{U_T} & \frac{1}{W} \frac{\partial W}{\partial U_{CB'}} \end{pmatrix} \qquad \text{(Gl. 6.68b)}$$

Dies ergibt mit den bisher benützten Werten (z. B. OC304), d. h. für $I_E = 1$ mA, $U_T = 25$ mV, $W = 40\,\mu$ und $\partial W/\partial U_{CB'} = 0{,}33\,\mu/\mathrm{V}$ die folgenden Werte

$$\left(Y'\right) = \begin{pmatrix} 4 \cdot 10^{-2} & -8 \cdot 10^{-6} \\ -4 \cdot 10^{-2} & 8 \cdot 10^{-6} \end{pmatrix} \Omega^{-1}$$

Zum Vergleich seien die aus dem Datenblatt der Type OC 304 für $I_E = 1$ mA, $U_{CB} = 5$ V und $f = 1$ kHz errechneten y-Werte angegeben:

$$\left(Y'\right) = \begin{pmatrix} 4{,}00 \cdot 10^{-2} & -9{,}02 \cdot 10^{-6} \\ -3{,}94 \cdot 10^{-2} & 9{,}40 \cdot 10^{-6} \end{pmatrix} \Omega^{-1}$$

Die gute Übereinstimmung zwischen gemessenen und gerechneten Werten zeigt die Brauchbarkeit der obigen Näherung. Wir sehen auch,

daß $$y'_{11}, y'_{21} \gg y'_{12}, y'_{22},$$

eine Tatsache, die wir uns für spätere Rechnungen merken wollen.

Weiterhin zeigt sich, daß $y'_{11} \approx y'_{21}$ und $y'_{12} \approx y'_{22}$, daß die y-Werte also sehr genau gemessen werden müssen, um unterschieden werden zu können. Dies ist eine Schwäche der Leitwertsmatrix, die dazu führt, daß der Schaltungstechniker bevorzugt mit der in Abschnitt 6.4.6 zu besprechenden h-Matrix arbeitet.

Wir können nun auch ein der Abb. 6.9 entsprechendes Wechselstromersatzschaltbild des inneren Transistors angeben (Abb. 6.15), das zwei Leitwerte, y'_{11} und y'_{22}, und zwei Stromgeneratoren, $y'_{12}\ u_{CB'}$ und $y'_{21}\ u_{EB'}$, enthält.

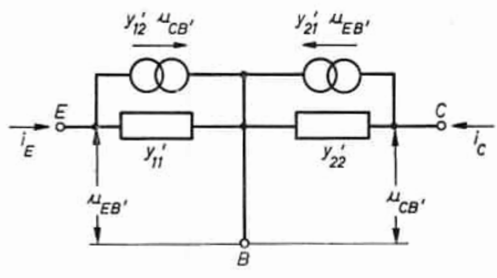


Abb. 6.15.
Ersatzschaltbild des „inneren" Transistors.

6.4.3 Der „äußere" Transistor mit Basiswiderstand und Feldzonenkapazitäten. Wie im Gleichstromfall (Abschnitt 6.3.3) kann der Wirkung des Basiswiderstandes durch Einsetzen eines Widerstandes $r_{BB'}$ entsprechender Größe in die Basiszuleitung des Ersatzschaltbildes Abb. 6.15 Rechnung getragen werden. Außerdem ist es notwendig, die Kapazitäten der PN-Übergänge zu berücksichtigen. Wir haben in den Abschnitten 5.4.2.4 und 5.6.2.1 gesehen, daß ein PN-Übergang zweierlei Arten von Kapazitäten besitzt. Die eine, die in Abschnitt 5.6.2.1 eingeführte Diffusionskapazität, ist durch die vollständige Lösung des Diffusionsproblems in Abschnitt 6.4.1 und 6.4.2 schon berücksichtigt. Noch nicht berücksichtigt ist jedoch die in Abschnitt 5.4.2.4 und in Gl. 5.32 definierte Feldzonen- oder auch Raumladungskapazität der beiden PN-Übergänge.

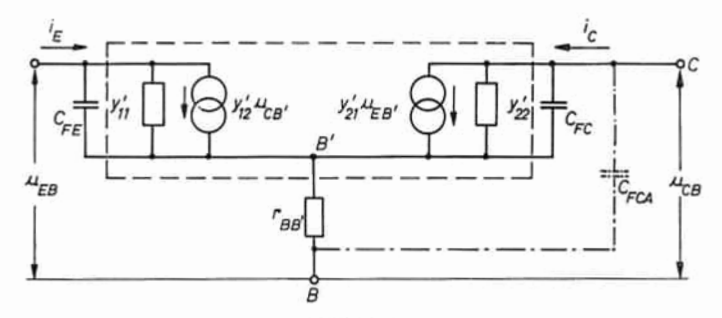


Abb. 6.16.
Ersatzschaltbild des „äußeren" Transistors.

Abb. 6.16 zeigt das in dieser Weise ergänzte Ersatzschaltbild. Die beiden Raumladungskapazitäten C_{FE} und C_{FC} sind direkt parallel zum Eingang bzw. Ausgang des inneren Transistors (gestrichelter Kasten) gezeichnet. Das ist streng genommen für einen Legierungstransistor mit unsymmetrischer Geometrie oder für einen Mesatransistor (Abb. 5.20) nicht ganz richtig, da der von der größeren Kollektorfläche kommende Verschiebungsstrom nach Abb. 6.10c

zum Teil über einen kleineren Basiswiderstand abfließen kann. Die äußere Kollektorkapazität C_{FCA}, die diesen Sachverhalt beschreibt, ist in Abb. 6.16 strichpunktiert angedeutet. Wir wollen darauf hier jedoch nicht näher eingehen, sondern zunächst mit der einfacheren ausgezogenen Schaltung rechnen.

Es ist wieder analog zu Gl. 6.34, 6.35

$$u_{EB} = u_{EB'} + r_{BB'}(i_E + i_C) \qquad \text{(Gl. 6.71)}$$

$$u_{CB} = u_{CB'} + r_{BB'}(i_E + i_C) \qquad \text{(Gl. 6.72)}$$

Die Vierpolkoeffizienten des gesamten äußeren Transistors, d. h. die Leitwerte der Gleichungen

$$\begin{aligned} i_E &= y_{11}\, u_{EB} + y_{12}\, u_{CB} \\ i_C &= y_{21}\, u_{EB} + y_{22}\, u_{CB} \end{aligned} \qquad \text{(Gl. 6.73)}$$

sind aus Abb. 6.16 in elementarer, aber etwas langwieriger Rechnung als Funktion der bekannten $y'_{11}, y'_{12}, \ldots$, der Kapazitäten C_{FE} und C_{FC} und von $r_{BB'}$ zu gewinnen. Man erhält

$$y_{11} = \frac{y'_{11} + i\,\omega\, C_{FE} + [r_{BB'}\, \Delta_y']}{1 + r_{BB'}\, \Sigma_y'} \qquad \text{(Gl. 6.74)}$$

$$y_{12} = \frac{y'_{12} - r_{BB'}\, \Delta_y'}{1 + r_{BB'}\, \Sigma_y'} \qquad \text{(Gl. 6.75)}$$

$$y_{21} = \frac{y'_{21} - [r_{BB'}\, \Delta_y']}{1 + r_{BB'}\, \Sigma_y'} \qquad \text{(Gl. 6.76)}$$

$$y_{22} = \frac{y'_{22} + i\,\omega\, C_{FC} + r_{BB'}\, \Delta_y'}{1 + r_{BB'}\, \Sigma_y'} \qquad \text{(Gl. 6.77)}$$

Hierbei ist:

$$\Delta_y' \equiv (y'_{11} + i\,\omega\, C_{FE})\,(y'_{22} + i\,\omega\, C_{FC}) - y'_{12}\, y'_{21} \qquad \text{(Gl. 6.78)}$$

$$\Sigma_y' \equiv y'_{11} + [y'_{12}] + y'_{21} + [y'_{22}] + i\,\omega\, C_{FE} + [i\,\omega\, C_{FC}] \qquad \text{(Gl. 6.79)}$$

Durch diese etwas formale Behandlung ist das Wechselstromverhalten des Transistors vollständig beschrieben. Wir wollen aber für die praktische Verwendung die obigen Gleichungen noch vereinfachen. Wegen der Kleinheit des EARLY-Effektes, d. h. wegen Gl. 6.68b mit den dazugehörigen Zahlenwerten, ist y'_{12} und $y'_{22} \ll y'_{11}$ und y'_{21}. Ebenso ist bei nicht zu hohen Frequenzen der Leitwert der Kollektor-Raumladungskapazität klein, also $\omega\, C_{FC} \ll y'_{11}, y'_{21}$. $1/r_{BB'}$ ist bei praktischen Transistoren von der Größenordnung y'_{11}. Wir können also annehmen, daß

$$y'_{22}, y'_{12}, \omega\, C_{FC} \ll y'_{11}, y'_{21}, 1/r_{BB'} \qquad \text{(Gl. 6.80)}$$

und damit

$$r_{BB'}\Delta_y' = (y'_{11} + i\omega C_{FE})\frac{y'_{22} + i\omega C_{FC}}{1/r_{BB'}} - y'_{21}\frac{y'_{12}}{1/r_{BB'}} \ll y'_{11}, y'_{21} \quad \text{(Gl. 6.81)}$$

Dadurch können wir in den Gleichungen 6.74 bis 6.79 einige Glieder vernachlässigen. Wir haben diese Glieder dort in eckige Klammern [ ] gesetzt und werden Gl. 6.74 bis 6.79 von jetzt ab ohne diese Glieder verwenden. Zum Zwecke der schnellen Übersicht seien sie noch einmal in Matrixform angeschrieben:

$$\begin{pmatrix} Y \end{pmatrix} = \frac{1}{1 + r_{BB'}\Sigma_y'}\begin{pmatrix} y'_{11} + i\omega C_{FE} & y'_{12} - r_{BB'}\Delta_y' \\ y'_{21} & y'_{22} + i\omega C_{FC} + r_{BB'}\Delta_y' \end{pmatrix} \quad \text{(Gl. 6.74–77)}$$

6.4.4 Die Verstärkungsfaktoren in Basisschaltung

6.4.4.1 Die Stromverstärkung: Die Stromverstärkung für Wechselstrom ist definiert als der Quotient aus Kollektor- und Emitterstrom. Sie ist im allgemeinen abhängig vom Lastwiderstand R_L. Man erhält sie aus den Vierpolgleichungen und dem Spannungsabfall am Lastwiderstand.

$$u_{CB} = -R_L i_C \quad \text{(Gl. 6.82)}$$

Wird dies in die beiden Gleichungen 6.73 eingesetzt und u_{EB} eliminiert, so erhält man die Stromverstärkung

$$V_I = \frac{i_C}{i_E} = \frac{y_{21}}{y_{11} + \Delta_y R_L} \quad \text{(Gl. 6.83)}$$

wobei

$$\Delta_y = y_{11}y_{22} - y_{12}y_{21} \quad \text{(Gl. 6.84)}$$

ist.

Daraus ergibt sich für $R_L = 0$ die Kurzschlußstromverstärkung, deren Betrag allgemein mit α bezeichnet wird und die dem Vorzeichen nach negativ ist.

$$\left(\frac{i_C}{i_E}\right)_{u_{CB}=0} = \frac{y_{21}}{y_{11}} \equiv -\alpha \quad \text{(Gl. 6.85)}$$

Mit den Gl. 6.74, 6.76, 6.64, 6.65, 6.58, 6.59 und der Näherung Gl. 6.68b für y'_{11} ergibt sich daraus

$$\alpha = -\frac{y'_{21}}{y'_{11}} \cdot \frac{1}{1 + i\omega C_{FE}/y'_{11}} = -\frac{\tilde{a}_{21}}{\tilde{a}_{11}} \cdot \frac{1}{1 + i\omega C_{FE}/y'_{11}}$$

$$= \frac{1}{\cosh\left(\dfrac{\eta W}{L_p}\right)} \cdot \frac{1}{1 + \dfrac{D_n L_p n_{PE}}{D_p L_{nE} p_N}\tanh\left(\dfrac{\eta W}{L_p}\right)} \cdot \frac{1}{1 + i\omega C_{FE} U_T / I_E} \quad \text{(Gl. 6.86)}$$

Das mittlere Glied ist die Wechselstrom-Emitterergiebigkeit $\tilde{\gamma}$ auf Grund des Elektronenstromes. Für den praktisch wichtigen Fall, wo $n_{PE} \ll p_N$, wird es zu Eins, und wir können weiter schreiben

$$\alpha \approx \frac{1}{\cosh\left(\frac{\eta W}{L_p}\right)} \cdot \frac{1}{1 + i\omega C_{FE} U_T/I_E} \qquad \text{(Gl. 6.87)}$$

Hier ist der erste Faktor der Wechselstromtransportfaktor oder die Stromverstärkung des inneren Transistors α' (mit $\tilde{\gamma} = 1$).

Der zweite Faktor kann als Hochfrequenz-Emitterwirkungsgrad aufgefaßt werden, der aber jetzt nicht durch den Elektronenanteil am Emitterstrom, sondern durch den Verschiebungsstrom der Emitter-Raumladungskapazität hervorgerufen wird. Wie man sieht, steigt er mit zunehmendem Strom.

Der erste Faktor α' wird für kleine W/L_p

$$\alpha' = \frac{1}{\cosh\left(\frac{\eta W}{L_p}\right)} \approx \frac{1}{1 + \frac{\eta^2 W^2}{2 L_p^2}} \approx \frac{1}{1 + i\omega \frac{W^2}{2 D_p}} \qquad \text{(Gl. 6.88)}$$

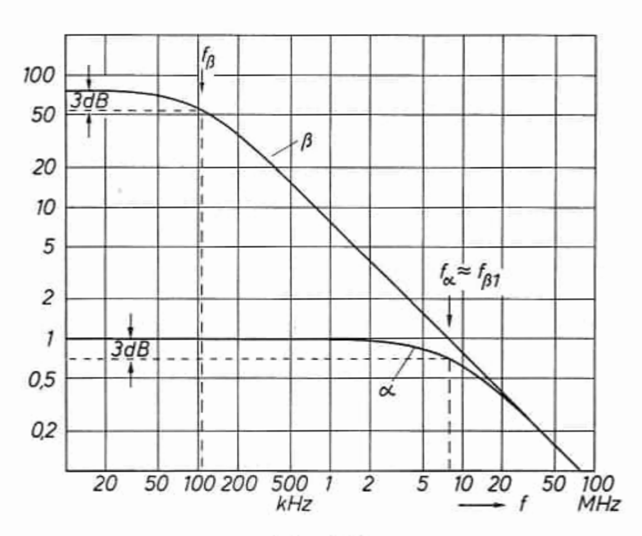


Abb. 6.17.
Frequenzverlauf der Stromverstärkungsfaktoren α und β eines Transistors.

Hier ist τ_p herausgefallen, weil $W \ll L_p$ vorausgesetzt wurde. Dieser Fall entspricht Abschnitt 5.6.2.2. Gl. 6.88 ergibt einen Abfall des Betrages von α' und damit von α mit steigender Frequenz

$$|\alpha'| = \frac{1}{\sqrt{1 + (\omega W^2/2\,D_p)^2}} \qquad \text{(Gl. 6.89)}$$

die in Abb. 6.17 für einen Transistor vom Typ OC 400 als Beispiel angegeben ist. Wir definieren die Grenzfrequenz f_L („Laufzeit"-Grenzfrequenz) als die Frequenz, bei der $|\alpha'|$ auf $1/\sqrt{2}$ seines Niederfrequenzwertes α_0 abgefallen ist. Mit Gl. 6.89 ergibt sich damit

$$f_L \equiv \frac{\omega_L}{2\pi} = \frac{D_p}{\pi W^2} \qquad \text{(Gl. 6.90)}$$

Diese Definition ist schon bei Gl. 6.61 benützt worden. Wird in Gl. 6.88 statt der Näherung der exakte Wert $1/\cosh\left(\frac{\eta W}{L_p}\right)$ benützt, so ergibt sich

$$f_L = 1{,}22\,\frac{D_p}{\pi W^2} \qquad \text{(Gl. 6.90a)}$$

Die Grenzfrequenz f_L nach Gl. 6.90 kann auch als Funktion der Laufzeit τ_C der Löcher durch die Basis aufgefaßt werden. Es ist nämlich

$$f_L = \frac{1}{2\pi\tau_C} \qquad \text{(Gl. 6.90b)}$$

wobei die Laufzeit

$$\tau_C = \frac{W^2}{2D_p} \qquad \text{(Gl. 6.91)}$$

st. Diese Gleichung läßt sich sehr einfach ableiten. Es sei Q_B die im stationären Zustand in der Basis vorhandene Gesamtladung der Löcher:

$$Q_B = e\,F\int_0^W p(x)\,dx \approx e\,F\,p_N\,e^{U_{EB'}/U_T} \cdot \frac{W}{2} \qquad \text{(Gl. 6.92)}$$

Zur Ausrechnung des Integrals wurde die für $W \ll L_p$ gültige Gl. 6.9 benützt) Diese Ladung ist innerhalb der Laufzeit τ_C gerade einmal aus dem Emitte. heraus und in den Kollektor hineingeflossen (bis auf Rekombinationsverluster.

Es ist deshalb

$$I_{Ep} = \frac{Q_B}{\tau_C}$$

oder

$$\tau_C = \frac{Q_B}{I_{Ep}} \qquad \text{(Gl. 6.93)}$$

Wenn darin Gl. 6.92 und Gl. 6.8a eingesetzt werden, gelangt man zu Gl. 6.91.

Der in Gl. 6.93 ausgedrückte einfache Zusammenhang zwischen Emitter- bzw. Kollektorstrom und Basisladung kann zu einer vereinfachten Beschreibung des Transistors, dem sogenannten Ladungssteuerungs-Prinzip, verwendet werden. Dieses Prinzip läßt sich mit Vorteil zum Vergleich des Transistors mit anderen Bauelementen und zur Betrachtung von Schaltvorgängen heranziehen. Wir können jedoch im Rahmen dieses Buches nicht weiter darauf eingehen und wollen auf die Literatur [91] verweisen.

Bei manchen Transistorstrukturen spielen noch andere Laufzeiten bei sehr hohen Frequenzen eine Rolle, z. B. die Laufzeit der Löcher durch die Kollektor-Raumladungszone. Diese Zeit ist dann in Gl. 6.90b zu τ_C zuzuzählen, um zum richtigen f_L zu gelangen.

Auch der zweite Faktor der Gl. 6.87, der Hochfrequenz-Emitterwirkungsgrad, fällt mit zunehmender Frequenz ab. Wir definieren eine Grenzfrequenz f_E (Emittergrenzfrequenz) als diejenige Frequenz, bei der dieser Faktor dem Betrag nach auf $1/\sqrt{2}$ seines Niederfrequenzwertes abgefallen ist. Dann ist analog zu Gl. 6.89 und 6.90

$$f_E \equiv \frac{\omega_E}{2\pi} = \frac{y'_{11}}{2\pi C_{FE}} \approx \frac{I_E}{2\pi C_{FE} U_T} \qquad \text{(Gl. 6.94)}$$

wobei diese Grenzfrequenz mit zunehmendem Strom steigt. Entsprechend läßt sich für das gesamte α eine Grenzfrequenz f_α definieren, durch

$$\alpha \equiv \frac{\alpha_0}{1 + i f/f_\alpha} \approx \frac{1}{1 + i f/f_\alpha} \qquad \text{(Gl. 6.95)}$$

Man erhält dann für den α-Verlauf bei nicht zu hohen Frequenzen (Vernachlässigung der Glieder mit ω^4)

$$\frac{1}{f_\alpha} = \frac{1}{f_L} + \frac{1}{f_E} \qquad \text{(Gl. 6.95a)}$$

Eine Diskussion dieser Größen erfolgt an Hand von Hochfrequenztransistoren in Abschnitt 6.5.

6.4.4.2 Spannungsverstärkung und Eingangsleitwert: Eine weitere wichtige Größe, die Spannungsverstärkung V_U, kann aus Gl. 6.73 gewonnen werden, wenn für

$$i_C = -y_L u_{CB} \qquad \text{(Gl. 6.96)}$$

eingesetzt wird, wobei y_L die (eventuell komplexe) Lastadmittanz ist. Man erhält dann

$$V_U = \frac{u_{CB}}{u_{EB}} = -\frac{y_{21}}{y_{22} + y_L} \qquad \text{(Gl. 6.97)}$$

Für Leerlauf am Ausgang ergibt sich damit $V_U = \alpha\, y_{11} / y_{22}$.

Mit Hilfe von Gl. 6.97 kann aus der ersten der Gleichungen 6.73 die Größe u_{CB} eliminiert werden und der Eingangsleitwert y_E gewonnen werden

$$y_E = y_{11} - \frac{y_{12}\, y_{21}}{y_{22} + y_L} \qquad \text{(Gl. 6.98)}$$

6.4.4.3 Die Leistungsverstärkung: Die Leistungsverstärkung kann man definieren als den Quotienten der Wirkleistung, die der Last zugeführt wird, und der Wirkleistung, die an der Eingangsseite dem Transistor zufließt.

$$\overline{V}_L = \frac{\frac{1}{2}\,\Re\mathrm{e}\,(u_{CB}\, i_C^*)}{\frac{1}{2}\,\Re\mathrm{e}\,(u_{EB}\, i_E^*)} = \left|\frac{u_{CB}}{u_{EB}}\right|^2 \cdot \frac{g_L}{g_E} \qquad \text{(Gl. 6.99)}$$

Hierbei sind i_C^* und i_E^* die konjugiert-komplexen Werte von i_C und i_E, g_L und g_E die Realwerte der Lastadmittanz bzw. des Eingangsleitwerts, $\Re\mathrm{e}\,(y_L) = \frac{1}{r_L}$ bzw. $\Re\mathrm{e}\,(y_E)$. Durch Einsetzen von Gl. 6.97 und 6.98 ergibt sich

$$\overline{V}_L = \left|\frac{y_{21}}{y_{22} + y_L}\right|^2 \frac{g_L}{\Re\mathrm{e}\left(y_{11} - \frac{y_{12}\, y_{21}}{y_{22} + y_L}\right)} \qquad \text{(Gl. 6.100)}$$

Diese Leistungsverstärkung wurde auf die tatsächlich dem Transistor zugeführte Eingangsleistung bezogen. Bei einem beliebig vorgegebenen Generator ist diese Leistung aber kleiner als diejenige, die bei optimaler Anpassung vom Generator verfügbar ist. Letztere ist nämlich (vgl. Gl. 5.104)

$$N_{G\max} = \frac{1}{4}\frac{|i_{G\,\mathrm{eff}}|^2}{g_G} = \frac{1}{8}\frac{|i_G|^2}{g_G} \qquad \text{(Gl. 6.101)}$$

wo g_G der Realteil der dem Stromgenerator i_G parallel gedachten Generatoradmittanz y_G ist. i_G ist die Stromamplitude, also gleich $\sqrt{2}\, i_{G_{\text{eff}}}$. Bezieht man die Leistungsverstärkung auf diese optimal verfügbare Generatorleistung, so erhält man eine kleinere Leistungsverstärkung $V_L < \overline{V}_L$, die im Englischen als „Transducer Gain“ bezeichnet wird (vgl. Gl. 5.104)

$$V_L = \frac{\frac{1}{2}\,\Re\mathrm{e}\,(u_{CB}\, i_C{}^*)}{N_{G\max}} = 4\, g_G\, g_L \left|\frac{u_{CB}}{i_G}\right|^2 \qquad \text{(Gl. 6.102)}$$

Mit $i_G = u_{EB}\,(y_E + y_G)$ und den Gleichungen 6.97 und 6.98 wird daraus

$$V_L = 4\, g_G\, g_L \left|\frac{y_{21}}{(y_{11} + y_G)\,(y_{22} + y_L) - y_{12}\, y_{21}}\right|^2 \qquad \text{(Gl. 6.102a)}$$

Wird der Generator an den Transistor angepaßt ($y_G = y_E{}^*$), dann geht diese Gleichung in Gl. 6.100 über.

Aus Gl. 6.102a geht hervor, daß die Verstärkung unendlich wird, d. h. daß der Transistor zu schwingen anfängt, wenn

$$(y_{11} + y_G)\,(y_{22} + y_L) - y_{12}\, y_{21} \leq 0$$

Ein wichtiges Maß für die Güte eines Hochfrequenztransistors ist die Verstärkung bei optimaler Anpassung von Generator und Last und bei Neutralisation der Rückwirkung (Unilateralisation). Letzteres bedeutet, daß durch Hinzuschalten äußerer Schaltelemente passiver Art dafür gesorgt wird, daß die Rückwirkung verschwindet, d. h. daß der Parameter y_{12} für das Gesamtsystem (also Transistor + Zusatzglied) gleich Null wird.

Prinzipiell wird eine Unilateralisation erreicht durch Parallelschaltung eines Leitwerts der Größe $-y_{12}$ zwischen den Klemmen E und C des Vierpols. Wir wollen hier nicht darauf eingehen, wie ein solcher negativer Leitwert zu realisieren ist, sondern auf die Literatur, z. B. auf das Buch von W. Gärtner, verweisen. Der resultierende Gesamtvierpol hat dann die Leitwertsmatrix

$$\left(\overline{Y}\right) \equiv \begin{pmatrix} \overline{y_{11}} & \overline{y_{12}} \\ \overline{y_{21}} & \overline{y_{22}} \end{pmatrix} = \begin{pmatrix} y_{11} + y_{12} & y_{12} - y_{12} \\ y_{21} - y_{12} & y_{22} + y_{12} \end{pmatrix}$$

was bei Verwendung von Gl. 6.74 bis 6.77 mit Gl. 6.80, Gl. 6.81 und mit $y'_{22} - y'_{12} \ll i\,\omega\, C_{FC}$

$$\left(\overline{Y}\right) \approx \frac{1}{1 + r_{BB'}\, \Sigma_y'} \begin{pmatrix} y'_{11} + i\,\omega\, C_{FE} & 0 \\ y'_{21} & i\,\omega\, C_{FC} \end{pmatrix} \qquad \text{(Gl. 6.103)}$$

ergibt.

Außerdem soll, wie gesagt, die Last angepaßt, also gleich dem konjugiert komplexen Wert des Ausgangsleitwerts gemacht werden, der in diesem Falle gleich $\overline{y_{22}}^*$ ist. Es ist also $y_L = \overline{y_{22}}^*$. Dann wird aus Gl. 6.100 die „Unilaterale, optimale Leistungsverstärkung"

$$V_{L\,\text{opt}} = \frac{\left(\overline{y_{21}}\right)^2}{4\,\mathfrak{Re}\left(\overline{y_{11}}\right)\mathfrak{Re}\left(\overline{y_{22}}\right)} \qquad \text{(Gl. 6.104)}$$

Beim Einsetzen der $\overline{y}$-Parameter aus Gl. 6.103 in diese Gleichung kann man einen praktisch wichtigen Ausdruck gewinnen, der häufig als Maß für die Güte von Hochfrequenztransistoren angegeben wird.

Nach Gl. 6.79, 6.86 und 6.95 ist

$$1 + r_{BB'}\,\Sigma_{y'} = 1 + r_{BB'}\,y'_{21}\left(\frac{1}{\alpha} - 1\right) = 1 + i\,r_{BB'}\,y'_{21}\,f/f_\alpha \qquad \text{(Gl. 6.105)}$$

und damit

$$\mathfrak{Re}\left(\overline{y_{11}}\right) = \frac{\mathfrak{Re}\left[\frac{y'_{21}}{\alpha}\left(1 - r_{BB'}\,y'^{*}_{21}\,f/f_\alpha\right)\right]}{\left|1 + i\,r_{BB'}\,y'_{21}\,f/f_\alpha\right|^2} \qquad \text{(Gl. 6.106)}$$

Wir wollen uns nun auf kleine Basisdicken W und nicht zu hohe Frequenzen beschränken, d. h. auf den Fall, wo $\eta W/L_p$ klein und deshalb nach Gl. 6.68a $\mathfrak{Im}\,(y'_{21}) \ll \mathfrak{Re}\,(y'_{21})$ ist. Dann erhält man unter Verwendung von Gl. 6.95

$$\mathfrak{Re}\left(\overline{y_{11}}\right) = y'_{21} \cdot \frac{1 + r_{BB'}\,y'_{21}\,(f/f_\alpha)^2}{1 + r^2_{BB'}\,y'^2_{21}\,(f/f_\alpha)^2} \approx y'_{21} \qquad \text{(Gl. 6.107)}$$

Der letzte Teil dieser Gleichung gilt, wenn

$$(r_{BB'}\,y'_{21}\,f/f_\alpha)^2 \approx \left(r_{BB'}\,\frac{I_E}{U_T}\,\frac{f}{f_\alpha}\right)^2 \ll 1 \qquad \text{(Gl. 6.108)}$$

und

$$r_{BB'}\,y'_{21}\,(f/f_\alpha)^2 \approx r_{BB'}\,\frac{I_E}{U_T}\left(\frac{f}{f_\alpha}\right)^2 \ll 1$$

also für kleinen Basiswiderstand, kleinen Strom und nicht zu hohe Frequenz. Aus Gl. 6.103 und 6.105 erhält man:

$$\mathfrak{Re}\left(\overline{y_{22}}\right) = \frac{\mathfrak{Re}\left[i\,\omega\,C_{FC}\,(1 - i\,r_{BB'}\,y'_{21}\,f/f_\alpha)\right]}{\left|1 + r_{BB'}\,\Sigma_{y'}\right|^2}$$

$$= \frac{\omega\,C_{FC}\,r_{BB'}\,y'_{21}}{\left|1 + r_{BB'}\,\Sigma_{y'}\right|^2}\,\frac{f}{f_\alpha} \qquad \text{(Gl. 6.109)}$$

Durch Einsetzen von $\overline{y_{21}}$, $\Re e(\overline{y_{11}})$ und $\Re e(\overline{y_{22}})$ aus Gl. 6.103, 6.107 und 6.109 erhält man schließlich

$$V_{L\,\mathrm{opt}} = \frac{f_\alpha}{8\,\pi\, f^2\, r_{BB'}\, C_{FC}} \qquad \text{(Gl. 6.110)}$$

Diese Gleichung wurde abgeleitet mit Hilfe der Gl. 6.74 bis 6.77, also ohne die in Abb. 6.16 strichpunktiert gezeichnete Aufspaltung von C_{FC} in die innere und die äußere Kollektorkapazität C_{FCA}. Wird dies jedoch bei Transistoren mit unsymmetrischer Geometrie, z. B. Mesatransistoren, berücksichtigt, so wird

$$y_{22} = y_{22}(C_{FCI}) + i\,\omega\, C_{FCA}$$

Der in Gl. 6.104 und 6.110 erscheinende Realwert von y_{22} enthält also lediglich die innere, also die dem Emitter gegenüberliegende Kollektorkapazität C_{FCI}. D. h., in Gl. 6.110 stellt C_{FC} nur diesen inneren Anteil der Kollektorkapazität dar.

Gl. 6.110 gibt zusammen mit den Ausdrücken für f_α (Gl. 6.95a, 6.94 und 6.90a), für $r_{BB'}$ (Gl. 6.43) und für C_{FC} (Gl. 5.32 und 5.30 bzw. 5.21a) im Prinzip eine genaue Vorschrift für die Dimensionierung eines Hochfrequenztransistors. Darauf soll in Abschnitt 6.5 ausführlicher eingegangen werden.

Ähnlich wie bei der Tunneldiode Gl. 5.107 kann man auch hier einen Gütefaktor definieren als das Produkt der Wurzel aus der Leistungsverstärkung und einer Frequenz. Wegen der speziellen Form von Gl. 6.110 kann eine beliebige Frequenz im Gültigkeitsbereich dieser Gleichung gewählt werden. Es ist dann

$$\sqrt{V_{L\,\mathrm{opt}}}\; f = \sqrt{\frac{f_\alpha}{8\,\pi\, r_{BB'}\, C_{FC}}} \equiv f_{\max} \qquad \text{(Gl. 6.110a)}$$

Diese Größe $f_{\max}$ wird auch als „Maximale Oszillationsfrequenz" bezeichnet Bei $f = f_{\max}$ ist nämlich gerade $V_{L\,\mathrm{opt}} = 1$.

6.4.5 Die Emitterschaltung. Wir haben bisher allen unseren Erörterungen stillschweigend die in Abb. 6.1 gezeichnete Schaltung zugrunde gelegt. Hier ist die Basisklemme dem Eingang und Ausgang gemeinsam und wird deshalb zweckmäßigerweise geerdet. Wir sprechen von der basisgeerdeten Schaltung. Außer dieser Schaltung gibt es noch zwei weitere Möglichkeiten, die Emitterschaltung und die Kollektorschaltung (siehe Abb. 6.18). Von allen drei Schaltungen ist die Emitterschaltung die am häufigsten verwendete und soll infolgedessen etwas eingehender behandelt werden.

Am Beispiel der Basisschaltung haben wir gesehen, daß das Wechselstromverhalten eines Vierpols grundsätzlich in allen Eigenschaften berechnet werden kann, wenn die Vierpolkoeffizienten, z. B. die y-Parameter, bekannt sind. Es genügt also, diese Parameter für die Emitterschaltung zu finden. Da es sich

aber hier um dasselbe Bauelement handelt wie in der Basisschaltung, müssen die Vierpolparameter für die Emitterschaltung aus denjenigen der Basisschaltung zu berechnen sein.

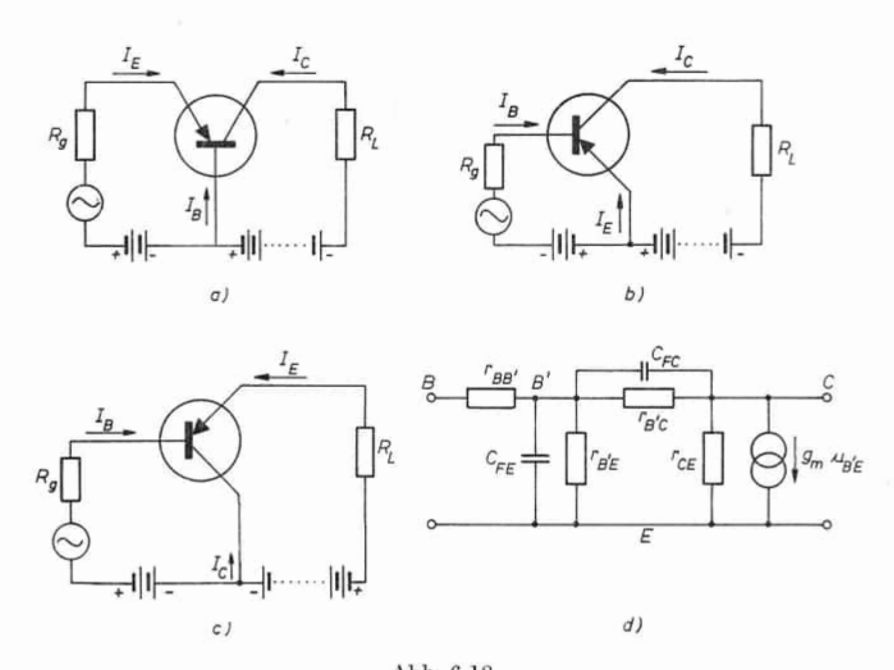


Abb. 6.18.
Die drei Schaltungsarten des Transistors (a, b, c) und das Hybrid-π-Ersatzschaltbild in Emitter-Schaltung (d).

Wir suchen die Koeffizienten y_{ikE} der Gleichungen zwischen Eingangs- und Ausgangsströmen und -spannungen

$$i_B = y_{11E}\, u_{BE} + y_{12E}\, u_{CE} \qquad \text{(Gl. 6.111)}$$

$$i_C = y_{21E}\, u_{BE} + y_{22E}\, u_{CE}$$

Die hier vorkommenden Ströme und Spannungen sind mit denen der entsprechenden Gleichungen der Basisschaltung Gl. 6.73 durch folgende Beziehungen verknüpft

$$i_B = - i_E - i_C \qquad \text{(Gl. 6.112)}$$

$$u_{EB} = - u_{BE} \qquad \text{(Gl. 6.113)}$$

$$u_{CB} = u_{CE} - u_{BE} \qquad \text{(Gl. 6.114)}$$

Werden Gl. 6.112 bis Gl. 6.114 und Gl. 6.73 in der Form der Gl. 6.111 geschrieben, so zeigt sich, daß

$$y_{11E} = y_{11B} + y_{12B} + y_{21B} + y_{22B} \qquad \text{(Gl. 6.115)}$$

$$y_{12E} = -(y_{12B} + y_{22B}) \qquad \text{(Gl. 6.116)}$$

$$y_{21E} = -(y_{21B} + y_{22B}) \qquad \text{(Gl. 6.117)}$$

$$y_{22E} = y_{22B} \qquad \text{(Gl. 6.118)}$$

ist, wobei sich die y_E auf die Emitter-, die y_B auf die Basisschaltung beziehen

Die Kurzschluß-Stromverstärkung in Emitterschaltung ist

$$\beta \equiv \left(\frac{i_C}{i_B}\right)_{u_{CE}=0} = \frac{y_{21E}}{y_{11E}} = -\frac{y_{21B} + y_{22B}}{y_{11B} + y_{12B} + y_{21B} + y_{22B}}$$

$$\approx \frac{-y_{21B}}{y_{11B} + y_{21B}} = \frac{\alpha}{1-\alpha} \qquad \text{(Gl. 6.119)}$$

Durchsichtiger ist eine andere Ableitung:

$$\beta = \left(\frac{i_C}{-i_C - i_E}\right)_{u_{CE}=0} = \left(\frac{-i_C/i_E}{1-(-i_C/i_E)}\right)_{u_{CE}=0}$$

$$\approx \left(\frac{-i_C/i_E}{1-(-i_C/i_E)}\right)_{u_{CB}=0} = \frac{\alpha}{1-\alpha} \qquad \text{(Gl. 6.120)}$$

Während α eine Zahl kleiner als Eins ist, kann β sehr groß werden. Je näher α an Eins herankommt, desto größer wird β. Die Emitterschaltung liefert also eine echte Stromverstärkung.

Man kann auch für β eine Grenzfrequenz f_β definieren, bei der $|\beta|$ auf $1/\sqrt{2}$ seines Niederfrequenzwertes β_0 abgefallen ist. Durch Einsetzen der Gl. 6.95 in Gl. 6.120 erhält man

$$\beta = \frac{\alpha_0}{1-\alpha_0} \cdot \frac{1}{1 + i\,\dfrac{f}{(1-\alpha_0)\,f_a}} = \frac{\beta_0}{1 + i \cdot \dfrac{f}{f_\beta}} \qquad \text{(Gl. 6.121)}$$

mit der β-Grenzfrequenz

$$f_\beta = (1-\alpha_0)\,f_a \approx \frac{f_a}{\beta_0} \qquad \text{(Gl. 6.122)}$$

Man sieht, daß β bei Gleichstrom zwar höher ist als α, aber dafür schon bei einer kleineren Frequenz abzufallen beginnt (vgl. Abb. 6.17). Tatsächlich wird

$|\beta| = 1$ bei einer Frequenz $f_{\beta 1}$, die (nach Gl. 6.120 bis Gl. 6.122) nahezu gleich der α-Grenzfrequenz ist

$$f_{\beta 1} = f_\beta \sqrt{\beta_0^2 - 1} \approx \beta_0 f_\beta \approx \frac{f_\beta}{1-\alpha_0} = f_\alpha \qquad \text{(Gl. 6.123)}$$

Ganz analog zu der Ableitung in Abschnitt 6.4.4.3 läßt sich die optimale Leistungsverstärkung $V_{L\,\mathrm{opt}}$ für die Emitterschaltung ableiten. Man geht von der für jeden Vierpol gültigen Gl. 6.104 aus, benützt die Beziehungen Gl. 6.115 bis Gl. 6.118 und gelangt mit entsprechenden Näherungen und für nicht zu tiefe Frequenzen wieder zu Gl. 6.110 bzw. Gl. 6.110a. Damit ist gezeigt, daß die Größe $f_{\max}$ als universaler Gütefaktor für beide Schaltungsarten brauchbar ist.

In Abb. 6.16 haben wir ein Ersatzschaltbild des Transistors in Basisschaltung angegeben. Es ergab sich rein formal aus den Leitwertsgleichungen des inneren Transistors und der Hinzufügung des Basiswiderstandes und der Raumladungskapazitäten. Dieses Ersatzschaltbild enthält zwei Ersatzstromquellen. Grundsätzlich kommt man aber mit *einer* Ersatzstromquelle oder -spannungsquelle aus. Derartige Ersatzschaltbilder sind in großer Zahl entwickelt worden (siehe z. B. das Buch von W. Gärtner).

Für die Emitterschaltung besonders wichtig und physikalisch sinnvoll ist das sogenannte Hybrid-π-Ersatzschaltbild von Giacoletto [93], das in Abb. 6.18d dargestellt ist. Hierin haben $r_{BB'}$, C_{FE} und C_{FC} die bisherige Bedeutung, während die vier anderen Elemente das Ersatzschaltbild des inneren Transistors darstellen. Wie man durch Vergleich mit den Leitwertsgleichungen des inneren Transistors, d. h. einer der Gl. 6.111 analogen Gleichung der gestrichenen Größen, leicht verifizieren kann, ist

$$r_{B'E} = \frac{1}{y'_{11E} + y'_{12E}}$$

$$r_{B'C} = -\frac{1}{y'_{12E}}$$

$$r_{CE} = \frac{1}{y'_{22E} + y'_{12E}}$$

$$g_m = y'_{21E} - y'_{12E} \qquad \text{(Gl. 6.124)}$$

Die Umwandlung in die $y_{B'}$-Werte mit Hilfe von Gl. 6.115 bis Gl. 6.118 und

die Anwendung von Gl. 6.68b ergibt

$$\frac{1}{r_{B'E}} = y'_{11B} + y'_{21B} = y'_{11B}(1 - \alpha') \approx \frac{I_E}{U_T}(1 - \alpha')$$

$$\frac{1}{r_{B'C}} = y'_{12B} + y'_{22B} = y'_{22B}(1 - \alpha'_I) \approx \frac{I_E}{U_T}(1 - \alpha'_I)$$

$$\frac{1}{r_{CE}} = - y'_{12B} \approx \frac{I_E}{W}\frac{\partial W}{\partial U_{CB'}} \qquad \text{(Gl. 6.125)}$$

$$g_m = - y'_{21B} + y'_{12B} \approx -y'_{21B} \approx \frac{I_E}{U_T}$$

Dabei wurden die Definitionen $\alpha' = y'_{21B}/y'_{11B}$ (Stromverstärkung des inneren Transistors) und $\alpha'_I = y'_{12B}/y'_{22B}$ (Stromverstärkung des inneren Transistors im inversen Betrieb) benützt. Mit den Zahlenwerten unseres bisher benützten Beispiels ($I_E = 1$ mA, $U_T = 25$ mV, $W = 40\,\mu$, $\partial W/\partial U_{CB'} = 0{,}33\,\mu/\text{V}$ und $\alpha' = 0{,}985$) und mit $\alpha'_I = 0{,}96$ erhält man

$$r_{B'E} = 1{,}67\ \text{k}\Omega$$

$$r_{B'C} = 3\ M\Omega$$

$$r_{CE} = 120\ \text{k}\Omega$$

$$g_m = 0{,}04\ \Omega^{-1}$$

6.4.6 Die h-Matrix. Wie in Abschnitt 6.4.2 schon erwähnt, hat die Leitwerts- oder y-Matrix den Nachteil, daß je zwei ihrer Koeffizienten nahezu gleich sind. Da ein aktiver Vierpol aber nur durch vier Parameter vollständig bestimmt ist, bedeutet dies, daß die y-Parameter sehr genau gemessen werden müssen. Aus diesem Grunde geht man gerne zur sogenannten Hybrid- oder h-Matrix über, die durch die folgenden Vierpolgleichungen für die Basisschaltung definiert wird.

$$u_{EB} = h_{11B}\, i_E + h_{12B}\, u_{CB} \qquad \text{(Gl. 6.126)}$$

$$i_C = h_{21B}\, i_E + h_{22B}\, u_{CB}$$

Wie man sieht, ist diese Matrix eine Mischung aus Impedanz-, Leitwerts- und dimensionslosen Elementen. Es ist nämlich

$$h_{11B} = \left(\frac{u_{EB}}{i_B}\right)_{u_{CB}=0}$$ Eingangsimpedanz bei kurzgeschlossenem Ausgang

$$h_{12B} = \left(\frac{u_{EB}}{u_{CB}}\right)_{i_E=0}$$ Spannungsrückwirkung bei offenem Eingang

$$h_{21B} = \left(\frac{i_C}{i_E}\right)_{u_{CB}=0}$$ Stromverstärkung bei kurzgeschlossenem Ausgang

$$h_{22B} = \left(\frac{i_C}{u_{CB}}\right)_{i_E=0}$$ Ausgangsleitwert bei offenem Eingang.

Die daraus resultierenden Meßvorschriften sind denkbar einfach, da es leichter ist, den hochohmigen Ausgang kurzzuschließen und den niederohmigen Eingang offen zu lassen, als umgekehrt. Die Verknüpfung der h-Parameter mit den y-Parametern ergibt sich, wenn die Gl. 6.73 in die Form Gl. 6.126 gebracht werden. Man erhält dann

$$\left(Y\right) = \begin{pmatrix} \frac{1}{h_{11}} & -\frac{h_{12}}{h_{11}} \\ \frac{h_{21}}{h_{11}} & \frac{h_{11}\,h_{22} - h_{12}\,h_{21}}{h_{11}} \end{pmatrix} \quad \text{(Gl. 6.127)}$$

und

$$\left(H\right) = \begin{pmatrix} \frac{1}{y_{11}} & -\frac{y_{12}}{y_{11}} \\ \frac{y_{21}}{y_{11}} & \frac{y_{11}\,y_{22} - y_{12}\,y_{21}}{y_{11}} \end{pmatrix} \quad \text{(Gl. 6.127a)}$$

Mit den y-Werten unseres Beispiels (OC 304) von Abschnitt 6.4.2 erhält man für die Basisschaltung

$$\left(H_B\right) = \begin{pmatrix} 25\,\Omega & 2{,}25 \cdot 10^{-4} \\ -0{,}985 & 5{,}25 \cdot 10^{-7}\,\Omega^{-1} \end{pmatrix}$$

Man sieht, daß sich diese h-Parameter alle um mindestens eine Größenordnung voneinander unterscheiden.

Die h-Matrix gelangt insbesondere für die Emitterschaltung zur Anwendung. Dafür lauten die Vierpolgleichungen

$$\begin{aligned} u_{BE} &= h_{11E}\; i_B + h_{12E}\; u_{CE} \\ i_C &= h_{21E}\; i_B + h_{22E}\; u_{CE} \end{aligned} \quad \text{(Gl. 6.126a)}$$

Durch Anwendung von Gl. 6.115 bis Gl. 6.118 und Gl. 6.127a können diese h_E-Parameter aus den y_B-Parametern errechnet werden. Für unser Beispiel (OC 304) erhält man

$$\left(H_E\right) = \begin{pmatrix} 1650\,\Omega & 6{,}5 \cdot 10^{-4} \\ 65 & 3{,}5 \cdot 10^{-5}\,\Omega^{-1} \end{pmatrix}$$

In der Emitterschaltung sind alle h-Parameter positiv und wieder alle von verschiedener Größenordnung.

Man kann nun auch die verschiedenen Verstärkungsfaktoren mit Hilfe der h-Parameter schreiben. So ist z. B. die Stromverstärkung in Emitterschaltung aus Gl. 6.83 durch Einsetzen der y_E- statt der y_B-Parameter und durch Verwendung von Gl. 6.127 zu gewinnen:

$$V_{IE} = \frac{y_{21E}}{y_{11E} + \Delta_{yE} R_L} = \frac{h_{21}}{1 + h_{22} R_L} \quad \text{(Gl. 6.128)}$$

woraus sich bei kurzgeschlossenem Ausgang ($R_L = 0$), $V_{IE} = h_{21E} \equiv \beta$ ergibt.

Als weiteres Beispiel sei die Leistungsverstärkung $\overline{V_{LE}}$ in der Emitterschaltung angeschrieben. Einsetzen der y_E-Parameter in Gl. 6.100 und Anwendung von Gl. 6.127 ergibt

$$V_{LE} = \left|\frac{y_{21E}}{y_{22E} + y_L}\right| \frac{g_L}{\Re e\left(y_{11E} - \dfrac{y_{12E}\, y_{21E}}{y_{22E} + y_L}\right)} = \frac{|h_{21E}|^2 R_L}{\Re e\,[(1 + h_{22E} R_L)(h_{11E} + \Delta_{hE} R_L)]}$$

Hierbei ist $\Delta_{hE} \equiv h_{11E}\, h_{22E} - h_{12E}\, h_{21E}$. (Gl. 6.129)

6.5 Der Transistor als Hochfrequenzverstärker

In den vorhergegangenen Abschnitten sind die grundlegenden Beziehungen schon abgeleitet worden, die für den Hochfrequenzverstärker von Bedeutung sind. Wir wollen diese Beziehungen im nächsten Abschnitt unter dem neuen Gesichtspunkt noch einmal betrachten und sie dann in den folgenden Abschnitten auf verschiedene für Hochfrequenzverstärkung in Frage kommende Transistortypen anwenden bzw. die Richtlinien für die Herstellung neuer Typen daraus ableiten.

6.5.1 Grundsätzliche theoretische Betrachtung. Als Ausgangspunkt wählen wir die optimale, neutralisierte Leistungsverstärkung $V_{L\,\mathrm{opt}}$ nach Gl. 6.110 oder die maximale Oszillationsfrequenz $f_{\max}$ nach Gl. 6.110a, wobei unser Bestreben sein muß, diese Gütefaktoren so groß wie möglich zu machen. Es muß also die Grenzfrequenz f_α möglichst groß, der Basiswiderstand $r_{BB'}$ und die innere Kollektorkapazität C_{FC} möglichst klein gemacht werden.

Die α-Grenzfrequenz ergibt sich aus Gl. 6.95a und Gl. 6.94 zu

$$\frac{1}{f_\alpha} = \frac{1}{f_L} + 2\pi \frac{U_T}{I_E} C_{FE} \qquad \text{(Gl. 6.130)}$$

Diese beiden die Grenzfrequenz bestimmenden Glieder sind bei verschiedenen Transistortypen verschieden groß. So wird z. B. bei Transistoren mit relativ großer Basisdicke W wegen Gl. 6.90 das erste Glied, also die Laufzeitgrenzfrequenz f_L, klein und wird bei nicht zu kleinen Emitterströmen I_E gegenüber der Emittergrenzfrequenz f_E bestimmend sein. Deshalb strebt man bei Hochfrequenztransistoren eine kleine Basisdicke W an. Andererseits ist gerade bei Transistoren für höchste Frequenzen der Stromstärke I_E eine Grenze gesetzt, wenn nachteilige Effekte auf die effektive Basisdicke und die Kollektorkapazität auf Grund einer Minoritäten-Überschwemmung des Kollektors vermieden

werden sollen [84]. Dementsprechend wird es dann wichtig, die Emitterfeldzonenkapazität C_{FE} klein zu machen. Es ist nach Gl. 5.32 und Gl. 5.30 für $N_D \ll N_A$ (guter Transistor, $\gamma \approx 1$)

$$C_{FE} = F_E \frac{\varepsilon \varepsilon_0}{l} = F_E \sqrt{\frac{\varepsilon \varepsilon_0 \, e N_D{}^E}{2(V_D - U_{EB'})}} \qquad \text{(Gl. 6.131)}$$

Danach müssen die Donatorendichte $N_D{}^E$ in der Basis am Emitterrand und die Emitterfläche F_E klein gemacht werden. Entsprechend kann die Forderung nach kleiner Kollektorkapazität C_{FC} durch kleine Donatorendichte $N_D{}^C$ am Kollektorrand und durch kleine wirksame Kollektorfläche erfüllt werden.

Die oben genannten Forderungen stehen aber in Widerspruch zu der Forderung, den Basiswiderstand klein zu machen, was z. B. nach Gl. 6.43 durch Vergrößern von N_D und W zu erfolgen hätte. Daraus folgt, daß bei der optimalen Dimensionierung eines Hochfrequenztransistors ein Kompromiß geschlossen werden muß, der wegen der verschieden großen Bedeutung der Glieder von f_α und $r_{BB'}$ bei den einzelnen Transistortypen unterschiedlich ausfallen wird. Dies soll in den folgenden Abschnitten an einigen Beispielen erläutert werden.

6.5.2 Der Legierungstransistor mit homogener Basis. Die Herstellung eines Legierungstransistors ist schon in Abschnitt 5.3.2.2 mit Abb. 5.17 besprochen worden. Dieser Prozeß liefert hohe Störstellenkonzentrationen im Emitter und Kollektor, eine sehr viel schwächer dotierte Basis und abrupte, unsymmetrische *PN*-Übergänge. Dies ist für einen *PNP*-Transistor in Abb. 6.19a dargestellt.

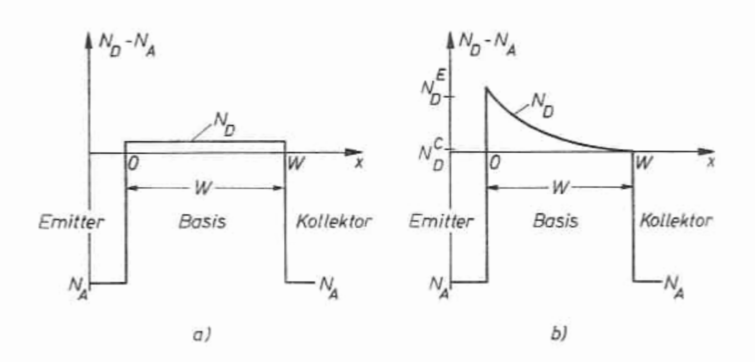


Abb. 6.19.
Störstoffverteilungen im normalen Transistor (a) und Drifttransistor (b)

Bei einem solchen Transistor ist die Basisdicke W noch relativ groß, so daß bei nicht zu kleinen Emitterströmen die Grenzfrequenz f_α durch das erste

Glied von Gl. 6.130, nämlich f_L, bestimmt wird. C_{FC} ist nach Abschnitt 6.4.4.3 die innere Kollektorkapazität, deren wirksame Fläche F_C mindestens zum Teil von der Emitterfläche F_E bestimmt wird (siehe auch Abschnitt 6.3.3 Ende, Abb. 6.10 und Abschnitt 6.4.3). Für den abrupten Kollektorübergang des Legierungstransistors ist wegen $|U_{CB'}| \gg V_D$

$$C_{FC} = F_C \sqrt{\frac{\varepsilon \varepsilon_0 e N_D}{2 U_{CB'}}} \qquad \text{(Gl. 6.132)}$$

Der Basiswiderstand ist durch Gl. 6.43 gegeben. Um ihn klein zu machen, muß zunächst der Basiskontakt so nahe an den Emitterrand herangebracht werden, als dies technologisch möglich ist. Außerdem sollten W_1 und W_2 (Abb. 6.10) so groß wie möglich gemacht werden, d. h. relativ dickes Halbleiterplättchen und tiefes Einlegieren. Auch hier trifft man schließlich auf eine technologische Grenze. Aber auch wenn durch diese Maßnahmen die äußeren Basiswiderstände (die ersten beiden Glieder von Gl. 6.43) sehr klein gemacht werden, so bleibt doch schließlich der Querwiderstand der eigentlichen Basisschicht (das dritte Glied von Gl. 6.43) übrig:

$$r_{BB'} \rightarrow \frac{1}{8\pi \sigma_N W} = \frac{1}{8\pi e \mu_n N_D W} \qquad \text{(Gl. 6.133)}$$

Die optimale Leistungsverstärkung wird dann nach Gl. 4.54, 6.90, 6.110, 6.130, 6.132 und 6.133 für eine gegebene Kollektorspannung und gegebene Frequenz

$$V_{L\,\mathrm{opt}} \sim \frac{f_\alpha}{r_{BB'} C_{FC}} \sim \frac{\mu_p / W^2}{(1/\mu_n N_D W)\, F_C \sqrt{N_D}} \sim \frac{\mu_p \mu_n \sqrt{N_D}}{F_C W} \qquad \text{(Gl. 6.134)}$$

Aus dieser Gleichung geht hervor, daß für den legierten Hochfrequenztransistor die Basisdicke W klein gemacht werden muß, wobei die untere Grenze von W nicht nur durch die Technologie, sondern auch durch die Spannungsfestigkeit des Transistors bestimmt sein wird, was in Abschnitt 6.6 eingehender behandelt werden soll. Weiterhin sehen wir aus Gl. 6.134, daß auch die Fläche F_C klein sein muß. Diesem ist jedoch ebenfalls durch die Technologie und durch die Größe der zu verstärkenden Leistung eine Grenze gesetzt. Außerdem sehen wir, daß N_D, die Dotierung der Basis, hoch gewählt werden muß, und zwar so hoch, wie es die Forderung nach einem guten Emitterwirkungsgrad

$$N_D = n_N \ll p_{PE}$$

und nach einer gewissen Kollektordurchbruchspannung noch zuläßt. Schließlich zeigt Gl. 6.134, daß das Produkt beider Beweglichkeiten eingeht, daß also *NPN*- und *PNP*-Transistoren grundsätzlich gleichwertig sind und daß Halbleitermaterialen mit hohen Beweglichkeiten, also z. B. Germanium, dem Silizium für Hochfrequenztransistoren vorzuziehen sind.

Auf Grund dieser Verhältnisse gelangt man bei einem derartigen einfachen, legierten Germanium-Hochfrequenztransistor zu folgenden Konstruktionsdaten:

Basisdicke W:	$15\,\mu$
Basisdotierung N_D:	$5 \cdot 10^{14}\,\text{cm}^{-3}$
(Spezifischer Widerstand:	$3\,\Omega$cm, n-Typ)
Emitterfläche F_E:	$1{,}6 \cdot 10^{-3}\,\text{cm}^2$
(Emitterdurchmesser:	$450\,\mu$)
Grenzfrequenz f_α:	8 MHz
Basiswiderstand $r_{BB'}$:	$100\,\Omega$
Wirksamer Teil der Kollektorkapazität C_{FC}:	10 pF
Maximale Oszillationsfrequenz:	18 MHz

6.5.3 Drifttransistor und PNIP-Transistor. Eine wesentliche Verbesserung der Hochfrequenzeigenschaften des legierten Transistors kann durch eine Erfindung von H. Krömer [39] erreicht werden, wonach die Dotierung in der Basis nicht mehr homogen wie in Abb. 6.19a, sondern vom Emitter zum Kollektor hin abfallend wie in Abb. 6.19b gewählt wird. Wir werden sogleich zeigen, daß dadurch ein inneres Feld in der Basis entsteht, ein sogenanntes „Drift"-Feld, das die Minoritätsträger sehr viel schneller vom Emitter zum Kollektor zieht, als das durch Diffusion allein (bei homogener Basis) möglich ist. Dementsprechend nennt man ein solches Bauelement „Drifttransistor".

Ein solcher Dotierungsgradient in der Basis läßt sich sehr einfach durch Eindiffusion von Verunreinigungen in das Halbleiterplättchen erreichen, wie in Abschnitt 5.3.3 und Abb. 5.18 ausgeführt wurde. Wird der Diffusionsprozeß sowieso angewandt zur Erzeugung des Kollektor-PN-Überganges, wie z. B. beim Mesatransistor (Abschnitt 5.3.3.2), so ergibt sich dieses Driftfeld von selbst. Wir werden in Abschnitt 6.5.4 sehen, daß fast alle modernen Hochfrequenztransistoren dieses Driftprinzip ausnützen. Zunächst sei aber ein einfacher Legierungstransistor betrachtet, bei dem durch einen Diffusionsprozeß vor dem Anlegieren der Emitter- und Basispillen ein Driftfeld in der Basis geschaffen wurde.

Bei inhomogener Dotierung, z. B. mit Donatoren $N_D(x)$, ergibt sich auch eine inhomogene Elektronendichteverteilung $n(x)$. Diese hat eine Diffusion der beweglichen Elektronen in Richtung abnehmender Dichte zur Folge. Dadurch wird die Ladungsneutralität zwischen negativen Elektronen und positiven Donatoren ein klein wenig gestört, wodurch ein elektrisches Feld entsteht. Dieses Feld ist bei Gleichgewicht, also verschwindendem Strom, genau groß genug, um den eben erwähnten Diffusionsstrom durch einen entgegengesetzt gerichteten Feldstrom zu kompensieren. Es gelten also ganz ähnliche Überlegungen wie in Abschnitt 4.1.4, nur daß dort ein primär vorhandenes Feld

einen Dichtegradienten zur Folge hatte, während hier primär ein Dichtegradient vorliegt, der (durch Diffusion) ein inneres Feld verursacht. Deshalb gilt auch hier die Formel für die Elektronendichte bei Gleichgewicht (Gl. 4.43)

$$n(x,y,z) \sim e^{\varphi / U_T} \tag{Gl. 6.135}$$

Durch Differentiation nach x erhält man daraus die Feldstärke

$$F = -\frac{d\varphi}{dx} = -U_T \frac{1}{n} \frac{dn}{dx} \tag{Gl. 6.136}$$

Wir können annehmen, daß der Dichtegradient klein genug ist, um die Raumladungsneutralität nur wenig zu stören (für eine quantitative Abschätzung siehe SHOCKLEY [33]); der andere Extremfall eines starken Gradienten mit großer Raumladung ist als PN-Sperrschicht realisiert).

Dann ist $n \approx N_D$, und wir haben

$$F = -U_T \frac{1}{N_D} \frac{dN_D}{dx} \tag{Gl. 6.137}$$

Das Feld F wird durch einen kleinen Injektionsstrom von Löchern nicht wesentlich gestört und treibt diese in Richtung fallender Dotierung, also vom Emitter zum Kollektor. Dieser Drifteffekt ist optimal, wenn das Feld konstant ist. Daraus folgt mit Gl. 6.137 für die optimale Verteilung der Dotierung in der Basis eines Drifttransistors

$$N_D(x) = N_D(x=0)\; e^{-\frac{Fx}{U_T}} \tag{Gl. 6.138}$$

also eine Exponentialverteilung, die in der Praxis durch die Eindiffusion von Donatoratomen hinreichend gut angenähert werden kann.

Wird $x = W$ in Gl. 6.138 eingesetzt, so erhält man das Driftfeld F als Funktion der Donatordichten $N_D{}^E$ und $N_D{}^C$ am Emitter- bzw. Kollektorrand

$$F = \frac{U_T}{W} \ln \frac{N_D{}^E}{N_D{}^C} \tag{Gl. 6.139}$$

Damit läßt sich die Laufzeit τ_C der Löcher durch die Basis abschätzen. Wir verwenden Gl. 2.2, Gl. 6.139, Gl. 4.54 und Gl. 6.91 und erhalten

$$\tau_{C\,\mathrm{Drift}} = \frac{W}{v_{\mathrm{Drift}}} = \frac{W}{\mu_p F} = \frac{W^2}{D_p \ln (N_D{}^E / N_D{}^C)}$$

$$= \tau_{C\,\mathrm{Diffusion}} \cdot \frac{2}{\ln (N_D{}^E / N_D{}^C)} \tag{Gl. 6.140}$$

Die Donatorendichte am Emitter muß für einen guten Emitterwirkungsgrad (z. B. $\gamma = 0{,}98$, damit $\beta \geq 50$) entsprechend dem nach Gl. 6.27 Gesagten um den Faktor 50 unter der Akzeptorkonzentration im Emitter liegen, die ihrerseits kleiner als die Entartungskonzentration $N_C = 10^{19}\,\text{cm}^{-3}$ (Tab. 3, Abschnitt 4.1.3) sein muß, damit kein Tunneleffekt auftritt. Man kommt damit auf $N_D{}^E \approx 10^{17}\,\text{cm}^{-3}$. Andererseits muß die Donatorendichte am Kollektor $N_D{}^C > n_i$ sein, also (bei Germanium) $N_D{}^C \approx 3 \cdot 10^{13}\,\text{cm}^{-3}$. Dies ergibt nach Gl. 6.139 ein inneres Feld F von etwa 200 V/cm bei einer Basisbreite von $10\,\mu$ und mit Gl. 6.140 eine etwa viermal kleinere Laufzeit beim Drifttransistor als beim normalen Transistor (Diffusionstransistor). Für die α-Grenzfrequenz ergibt dies jedoch einen noch größeren Verbesserungsfaktor, da es hierbei nicht eigentlich auf die Phase, sondern auf die Amplitude der den Kollektor erreichenden Löcherwelle ankommt. Eine Wiederholung der Rechnung von Abschnitt 6.4.1 unter Berücksichtigung des eingebauten Driftfeldes ergibt [39]

$$f_{L\,\text{Drift}} = f_{L\,\text{Diffusion}} \left(\frac{1}{2} \ln \frac{N_D{}^E}{N_D{}^C} \right)^{3/2} \qquad \text{(Gl. 6.141)}$$

Dies gibt mit obigen Zahlen eine Verbesserung von einem Faktor $4^{3/2} = 8$. Normalerweise ist $N_D{}^E$, die Donatordichte am Emitter, beim Drifttransistor größer als beim normalen Transistor, was zu einer vergleichsweise höheren Emitter-Raumladungskapazität C_{FE} führt (Gl. 6.131). Außerdem muß der Drifttransistor mit relativ kleinen Stromdichten betrieben werden, weil sonst durch zu hohe Löcherdichte das Driftfeld „zugeweht“ wird. Aus diesem Grunde fällt hier das zweite Glied von f_α, die Emittergrenzfrequenz f_E, stärker ins Gewicht, was den obigen Vorteil bezüglich f_α etwas mindert.

Wichtiger aber ist ein anderer Vorteil des Drifttransistors, der sich auf die Kollektorkapazität C_{FC} und auf den Basiswiderstand $r_{BB'}$ auswirkt. Nach Gl. 6.132 ist $C_{FC} \sim \sqrt{N_D}$, d. h. genau genommen der Donatorendichte $N_D{}^C$ am Kollektor. Diese ist jetzt aber nicht mehr gleich $N_D{}^E$ wie beim normalen Transistor, sondern sehr viel niedriger, wodurch C_{FC} wesentlich kleiner wird, was direkt in die Leistungsverstärkung $V_{L\,\text{opt}}$ bzw. $f_{\max}$ eingeht. Weiterhin hat der niedrige Wert von $N_D{}^C$ zur Folge, daß die Durchbruchspannung des Kollektors wesentlich größer ist (vgl. Abschnitt 5.7.2). Die Forderung einer hohen Spannungsfestigkeit begrenzt also jetzt nicht mehr die Donatorendichte in der ganzen Basis. Man kann deshalb mit dem emitterseitigen Wert $N_D{}^E$ wirklich bis an die oben erwähnte Grenze, die durch den Emitterwirkungsgrad γ gegeben ist, herangehen. Dadurch wird insbesondere der Basiswiderstand $r_{BB'}$ kleiner, was $V_{L\,\text{opt}}$ und $f_{\max}$ weiter erhöht.

Dies wird sehr eindrucksvoll gezeigt durch den Vergleich der Daten des normalen HF-Transistors am Schluß von Abschnitt 6.5.2 mit den folgenden

Daten eines typischen Germanium-Drifttransistors:

Basisdicke W:	$10\,\mu$
Basisdotierung am Emitter $N_D{}^E$:	$3 \cdot 10^{16}\,\text{cm}^{-3}$
Emitterfläche F_E:	$3 \cdot 10^{-4}\,\text{cm}^2$
(Emitterdurchmesser:	$200\,\mu$)
Grenzfrequenz f_α:	80 MHz
Basiswiderstand $r_{BB'}$:	$15\,\Omega$
Wirksamer Teil der Kollektor-kapazität C_{FC}:	4 pF
Maximale Oszillationsfrequenz:	230 MHz

Die oben erwähnte Entkopplung der Kollektorkapazität und der Durchbruchsspannung des Kollektors von der Donatorendichte am Emitterrand, also vom Basiswiderstand, setzt an sich nicht eine exponentielle Donatorenverteilung voraus. Z. B. kann der Übergang von der hohen Donatorendichte $N_D{}^E$ an der Emitterseite zur niedrigen Donatorendichte $N_D{}^C$ (oder gar zur Eigenleitung) am Kollektor stufenförmig erfolgen. Dies ist die Grundlage des *PNIP*-Transistors von Early [38], wie er schon in Abb. 5.10c im Prinzip angedeutet worden ist. Abb. 6.20 gibt einen Vergleich der beiden Transistorarten, wobei

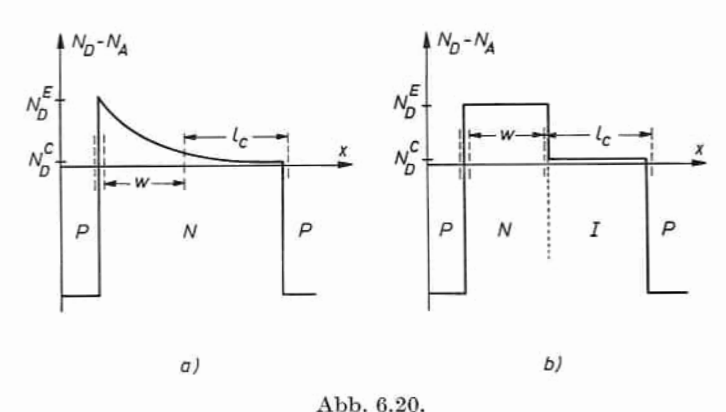


Abb. 6.20.
Störstoffverteilungen im Drifttransistor (a) und PNIP-Transistor (b)

Abb. 6.20a eine Wiederholung von Abb. 6.19b darstellt mit der Ausnahme, daß jetzt die Ausdehnung der Kollektorsperrschicht l_C bei hoher Sperrspannung eingezeichnet ist. Während wir bisher stillschweigend so getan haben, als sei W im wesentlichen der Abstand zwischen den beiden *PN*-Übergängen, sehen wir jetzt, daß die Sperrschichtausdehnung l_C im schwachdotierten rechten Teil der Basis beträchtlich ist. Darin liegt selbstverständlich die Ursache für die kleine Kapazität C_{FC} und die hohe Spannungsfestigkeit des Kollektors. In den obigen Formeln für τ_C und f_L ist dann natürlich die um l_C verkleinerte Größe W

zu benützen. Ein Blick auf Abb. 6.20b, wo der EARLYsche *PNIP*-Transistor dargestellt ist, zeigt die Ähnlichkeit der Verhältnisse.

Beim reinen *PNIP*-Transistor ist theoretisch allerdings kein Driftfeld vorhanden. Da aber praktisch eine solche Struktur immer durch Diffusionsprozesse hergestellt wird, entsteht automatisch ein vorteilhafter Gradient in der Basis. Somit sind beide Transistortypen, obwohl beide Erfindungen ursprünglich von verschiedenen Standpunkten ausgegangen sind, im wesentlichen identisch.

6.5.4 Diffundierte Höchstfrequenztransistoren. Es hat in den letzten Jahren nicht an originellen Versuchen und Entwicklungsanstrengungen gefehlt, auf allen möglichen Wegen *PNP*- bzw. *NPN*-Strukturen mit dünnen Basiszonen und kleinen Kapazitäten herzustellen. Als technisch überlegen hat sich aber eine Gruppe von Verfahren herausgestellt, die wir unter der obigen Überschrift zusammenfassen wollen. Es sollen dabei nur die Transistortypen eingeschlossen sein, bei denen die Basiszone selbst und damit der Kollektorübergang durch Diffusion von Störstellen erzeugt wird. Insbesondere handelt es sich hierbei um Mesatransistoren und Planartransistoren, wobei unter Umständen das sogenannte „Post-Alloy-Diffusion"-Prinzip und das Epitaxie-Verfahren mit verwendet werden.

Zunächst sollen die einzelnen in Frage kommenden technologischen Methoden ganz kurz erläutert und dann für alle diese Typen gemeinsam die Theorie der Dimensionierung angedeutet werden.

6.5.4.1 Der Mesatransistor: Die Herstellung eines Germanium-Mesatransistors [85] ist in Abschnitt 5.3.3.2 und Abb. 5.20 schon beschrieben worden. Wie dort ausgeführt worden ist, entsteht dabei der Kollektorübergang durch eine oberflächliche Diffusion von Störstellen in den Kollektorkörper hinein, während der Emitterübergang durch Aufdampfen und Einlegieren von geeigneten Metallflecken erzeugt wird. Als Basisdicke erhält man somit den Abstand zwischen der Diffusionstiefe und der Legierungsfront. Da der Legierungsprozeß stets etwas weniger genau beherrscht werden kann, ist es erstrebenswert, auch den Emitter durch Diffusion herzustellen. Das ist bei Germanium recht schwierig, da hier keine zur entsprechenden Diffusion geeigneten Stoffe zur Verfügung stehen und da insbesondere keine Maskierungsverfahren vorliegen.

Bei Silizium ist jedoch die bei der Oxydation an der Oberfläche entstehende Quarzhaut eine ideale Maske für die Emitterdiffusion. Der Herstellungsgang eines typischen Silizium-Mesatransistors ist in Abb. 6.21 dargestellt. In eine *N*-Typ-Siliziumplatte wird Bor eindiffundiert mit einer Eindringtiefe von wenigen μ und die Platte thermisch oxydiert. Dann wird eine photoempfindliche Lackschicht aufgebracht und durch entsprechende Blenden belichtet, wobei sich die belichteten Teile durch Polymerisation verfestigen. Eine nachfolgende

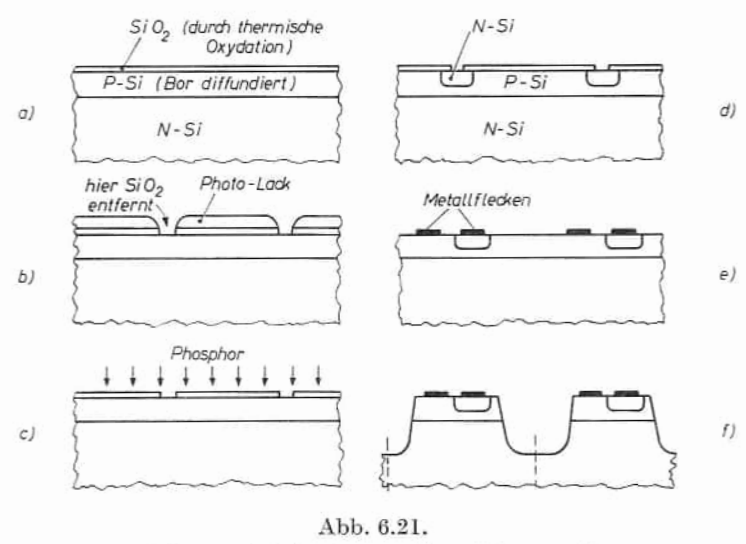


Abb. 6.21.
Herstellungsschritte eines Silizium-Mesatransistors.

Behandlung mit einem „Entwickler" löst die unbelichteten Lackteile heraus, so daß kleine Fenster im Lack entstehen, z. B. $200\,\mu \times 50\,\mu$. Durch eine Flußsäureätzung wird die Quarzschicht innerhalb dieser Fenster aufgelöst, während der stehengebliebene Photolack die anderen Teile schützt. Nach Abwaschen des Lackes dient die SiO_2-Schicht als Diffusionsmaske für die Herstellung der Emittoren (Phosphor-Diffusion). Schließlich wird die ganze SiO_2-Schicht abgeätzt, Metallflecken zur Kontaktierung von Emitter und Basis aufgebracht, die Mesaberge herausgeätzt, wie in Abschnitt 5.3.3.2 erwähnt, und die Siliziumplatte in Einzelelemente getrennt.

Auf diese Weise hat man die kritischen Größen, nämlich die Konzentration und Dicke der Basisschicht, ausschließlich durch Diffusionsschritte bestimmt, die sich heutzutage sehr genau beherrschen lassen.

6.5.4.2 Der Planartransistor: Beim oben beschriebenen Silizium-Mesatransistor wurde die durch thermische Oxydation auf dem Silizium erzeugte Quarzhaut als Diffusionsmaske benützt. Eine solche Maskierung kann nun weiterhin ausgenützt werden, um auch dem fertigen Transistor einen gewissen Oberflächenschutz zu verleihen und damit seine Stabilität zu erhöhen. Dies geschieht im sogenannten Planarverfahren [86], das in Abb. 6.22 angedeutet ist. Hier wird die Siliziumplatte vor der ersten Diffusion mit einem Quarzfilm versehen, der Fenster enthält. Die darauffolgende Diffusion bildet deshalb keine zusammenhängende Basisschicht, sondern Basisflecken, und zwar so, daß der Kollektor-*PN*-Übergang nicht an die freie Oberfläche stößt, sondern von der Quarzschicht geschützt ist. Bei den nachfolgenden Schritten wird wieder

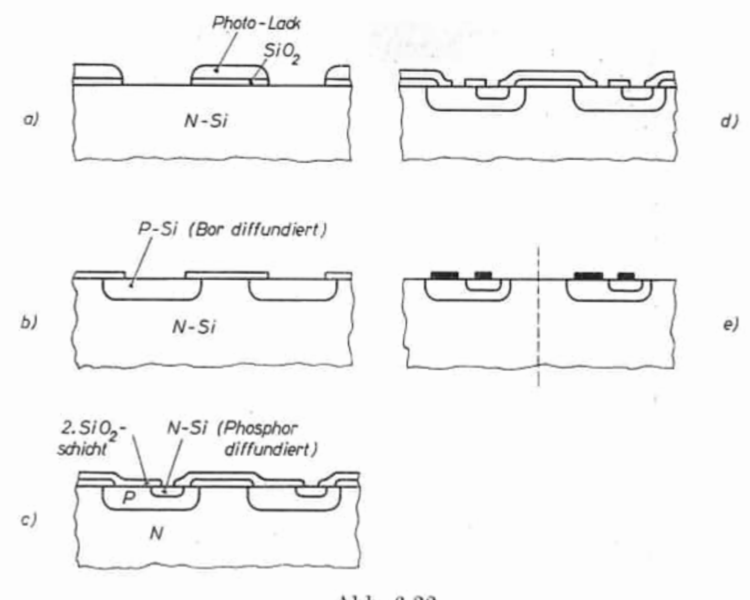
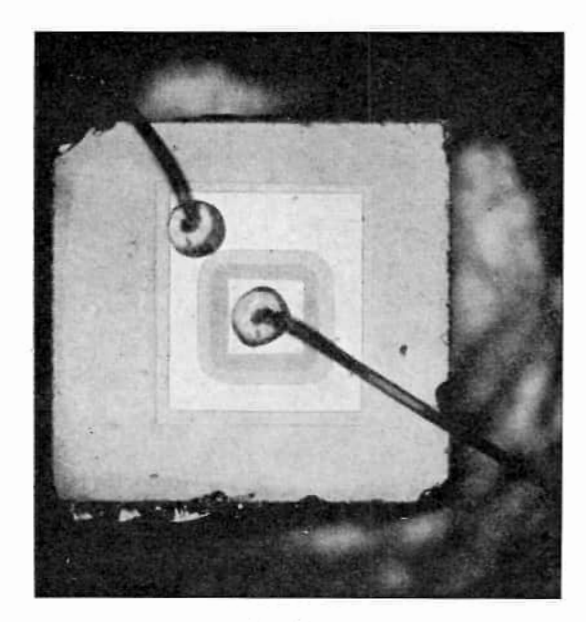


Abb. 6.22.
Herstellungsschritte eines Silizium-Planartransistors.

Oxydation und Diffusion von Phosphor durch die Fenster angewandt, um die Emittoren zu erzeugen. Die Oxydschichten bleiben aber jetzt erhalten; zum Zwecke der Kontaktierung werden lediglich wieder Fenster aus der Oxydhaut herausgeätzt. Damit hat man einen Transistor, dessen *PN*-Übergänge niemals an freie Oberflächen, sondern nur an Oberflächen anstoßen, die von einer vor dem Diffusionsvorgang natürlich gewachsenen Quarzschicht bedeckt sind. Infolgedessen erzielt man mit solchen Transistoren kleinere Leckströme und bessere Stabilität als mit jedem anderen Transistortyp. Einen Blick durch das Mikroskop auf einen solchen Transistor zeigt Tafel 17.

6.5.4.3 Das „Post-Alloy-Diffusion"-Prinzip: Wie der Name dieses Verfahrens [87] andeutet, wird hier nach der Bildung des Emitters durch Legieren die eigentliche Basisschicht durch Diffusion eines Störstoffes aus der legierungsdotierten Emitterschicht heraus erzeugt. Dies hat den Vorteil, daß die Basisdicke nur durch einen Diffusionsschritt bestimmt wird, unabhängig von der Legierungstiefe des Emitters und der Form der Legierungsfront. Ein Durchlegieren (also ein Kurzschluß) des Emitters durch die Basis ist ausgeschlossen.

Dieses Verfahren wird im allgemeinen in der Form eines Germanium-Mesatransistors mit pillenförmigen Emitter- und Basisanschlüssen realisiert. Es kann jedoch auch als Verfeinerung im oben erwähnten Mesa mit Aufdampfflecken verwendet werden. Die grundsätzlichen Schritte seien in Abb. 6.23 angedeutet. Zunächst wird eine Vordiffusion eines Donators in die *P*-Germanium-



Tafel 17.
Aufsicht auf einen Silizium-Planar-transistor (Werkphoto Intermetall).

Platte durchgeführt, um den Basisanschluß zu ermöglichen. Sodann werden zwei Pillen oder Flecken einlegiert: erstens eine Basispille, die neben einem Trägermetall (z. B. Blei) einen Donator enthält, wodurch ein ohmscher Kontakt mit der vordiffundierten Schicht hergestellt wird, und zweitens eine

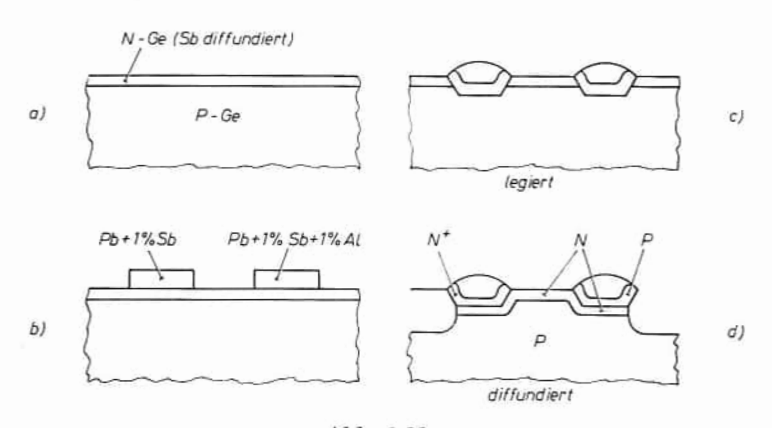


Abb. 6.23.
Herstellungsschritte eines „Post-Alloy-Diffused"-Transistors.

Emitterpille, die außerdem zusätzlich eine höhere Akzeptorkonzentration enthält. Bei richtiger Wahl dieses Stoffsystems baut sich der Akzeptor in größerer Konzentration in die Rekristallisationszone ein als der Donator. Dadurch entsteht die P-Emitterzone. In einem nachfolgenden Diffusionsschritt diffundiert der Donator aber schneller aus der Emitter-Rekristallisationszone in die Geplatte hinein als der Akzeptor. Es entsteht die eigentliche, sehr dünne Basiszone. Anschließend kann auch hier nach geeigneter Abdeckung mit Wachs ein Mesaberg geätzt werden.

6.5.4.4 Das Epitaxie-Verfahren: Die oben beschriebenen Transistorstrukturen haben gemeinsam, daß Basis und Emitter durch Eindiffundieren bzw. Einlegieren in eine relativ dicke Kollektorzone entstehen. Diese Kollektorzone macht auch beim fertigen Bauelement den größten Teil der Dicke der Halbleiterplatte aus. Um die Kollektorkapazität klein und die Kollektorabbruchspannung hoch zu halten, muß diese Kollektorplatte relativ schwach dotiert, also hochohmig sein. Daraus ergibt sich ein recht beachtlicher Kollektorserienwiderstand, der die Verstärker- und Schalteigenschaften des Transistors wesentlich verschlechtert.

Der Ausweg aus dieser Schwierigkeit ist (ähnlich wie beim $PNIP$-Prinzip für die Basis) eine inhomogene Störstellenverteilung im Kollektor. Insbesondere wünscht man eine dünne hochohmige Schicht in der Nähe des PN-Übergangs und den übrigen Teil der Platte niederohmig. Es ist sehr schwierig, eine solche Verteilung durch Diffusionsprozesse zu erreichen, da durch Diffusion grundsätzlich nur dünne niederohmige Schichten an hochohmigem Grundmaterial reproduzierbar erzeugt werden können und nicht das Umgekehrte. Man zieht deshalb hierzu das sogenannte Epitaxie-Verfahren [88] heran.

Dabei wird die niederohmige, einkristalline Silizium- oder Germaniumscheibe bei erhöhter Temperatur dem Dampf einer Silizium- bzw. Germaniumverbindung (z. B. $SiCl_4$, GeJ_4) ausgesetzt, die sich bei dieser Temperatur leicht in ihre Bestandteile zerlegen. Bei richtiger Führung dieses Zersetzungsprozesses schlägt sich auf der Halbleiterplatte eine hochohmige Germanium- bzw. Siliziumschicht ab, und zwar einkristallin, epitaxial, d. h. in Fortsetzung der einkristallinen Struktur der Grundplatte. Diese Epitaxieschichten können so gleichmäßig aufwachsen, daß die Platten ein hervorragendes Ausgangsmaterial für Mesa- und Planartransistoren darstellen. In Abb. 6.24 ist der schematische Aufbau eines solchen Planar-Epitaxial-Transistors dargestellt.

6.5.4.5 Dimensionierung: Wie ein vergleichender Blick auf die Abb. 6.21, 6.22, 6.23 und 6.24 zeigt, lassen sich alle diese „diffundierten Höchstfrequenztransistoren“ durch ein gemeinsames Modell behandeln. Wir wollen dafür das in Abb. 6.25 gezeichnete Modell eines PNP-Mesatransistors mit rechteckigem Emitter- und Basisstreifen verwenden. Die Abweichung, die durch die Krümmung des Kollektorüberganges beim Planartransistor und beim Post-Alloy-

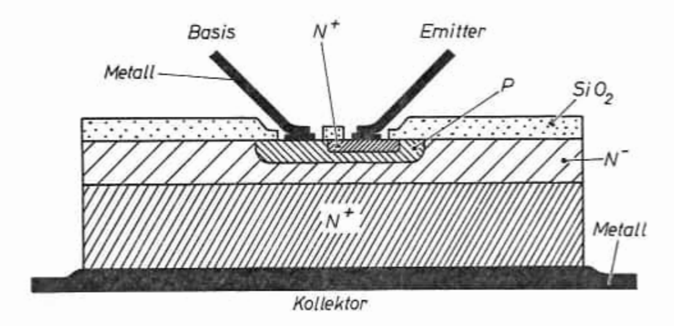


Abb. 6.24.
Querschnitt durch einen Epitaxie-Planartransistor.

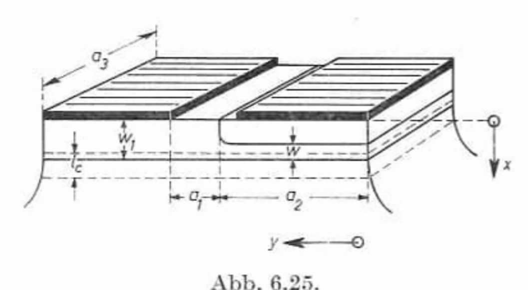


Abb. 6.25.
Zur Dimensionierung des diffundierten Höchstfrequenztransistors.

Diffused-Transistor auftritt, soll vernachlässigt werden. Ebenso ist die Umrechnung auf eine runde Emitterpille trivial. Außerdem soll zunächst der Einfluß des Kollektorserienwiderstandes vernachlässigt werden.

Da die Basisschicht durch Diffusion entstanden ist, besitzt sie ein Driftfeld. Sie ist deshalb auch wesentlich niederohmiger als der Kollektorkörper. Aus diesem Grunde bildet sich die Kollektorsperrschicht bevorzugt in diesen hinein aus, und wir können in unserer vereinfachten Analyse die effektive Basisbreite W als gleich dem Abstand vom Emitter- zum Kollektorübergang annehmen.

Zunächst ist die Grenzfrequenz f_α zu bestimmen. Wir verwenden Gl. 6.130 und berücksichtigen das Driftfeld mit Gl. 6.141. Wegen der hochohmigen Kollektorzone ist die Kollektorsperrschichtdicke l_C relativ groß (unter Umständen $l_C > W$), so daß wir die Laufzeit durch die Kollektorsperrschicht zu berücksichtigen haben. In dem sehr hohen Feld dieser Sperrschicht werden die Löcher so stark beschleunigt, daß sie dem ohmschen Gesetz nicht mehr folgen, sondern eine von der Feldstärke unabhängige Grenzgeschwindigkeit v_{lim} von etwa $6 \cdot 10^6$ cm/sec erreichen. Auf diese Abweichung vom ohmschen Gesetz bei

hohen Feldern ist schon in Abschnitt 2.3.1 hingewiesen worden. Dadurch wird die Rechnung für uns besonders einfach. Man hat für die der Laufzeit durch die Kollektorsperrschicht entsprechende Grenzfrequenz f_{LC}

$$f_{LC} = \frac{v_{\lim}}{\pi\, l_C} \qquad \text{(Gl. 6.142)}$$

[Wegen des im Vergleich zu Gl. 6.90b fehlenden Faktors 2 siehe z. B. das Buch von HUNTER (1962) S. 4–24]. Durch Addition der reziproken Grenzfrequenzen erhält man

$$\frac{1}{f_a} = \frac{1}{f_{LB}} + \frac{1}{f_{LC}} + \frac{1}{f_E} \qquad \text{(Gl. 6.143)}$$

$$= \frac{\pi W^2}{D_p}\left[\frac{2}{\ln(N_D{}^E / N_D{}^C)}\right]^{3/2} + \frac{\pi\, l_C}{v_{\lim}} + \frac{2\pi\, U_T\, C_{FE}}{I_E}$$

Es ist hier angebracht, einige Zahlenwerte für diese Größen anzuführen. Als typische Konstruktionsdaten für derartige Germaniumtransistoren sollen die folgenden Werte gewählt werden:

$$\begin{array}{ll}
W_1 = 3\,\mu & N_D(x=0) \equiv N_D{}^0 = 2{,}5 \cdot 10^{17}\ \mathrm{cm^{-3}} \\
W = 1{,}5\,\mu & N_D{}^E = 5 \cdot 10^{16}\ \mathrm{cm^{-3}} \\
a_1 = 25\,\mu & N_D{}^C = \mathrm{N}_A{}^C = 5 \cdot 10^{14}\ \mathrm{cm^{-3}} \\
a_2 = 50\,\mu & I_E = 1\ \mathrm{mA} \\
a_3 = 100\,\mu & U_{CB'} = 10\ \mathrm{V}
\end{array}$$

Daraus ergibt sich $f_{LB} = 2{,}4$ GHz. Unter Benutzung von Gl. 5.30 mit dem dortigen $N_D \gg N_A = N_D{}^C$, $U = U_{CB'}$ und $v_{\lim} = 6 \cdot 10^6$ cm/sec erhält man $f_{LC} = 3{,}2$ GHz. Mit Gl. 6.131 und mit der Abschätzung $V_D - U_{EB'} \approx 0{,}3$ V ergibt sich schließlich $f_E = 0{,}97$ GHz. Man sieht an diesem Beispiel, daß alle drei Glieder von Gl. 6.143 wichtig sind, f_E aber noch ziemlich dominierend ist. Bei einem Strom $I_E > 3$ mA beginnt jedoch f_E schon seinen Einfluß zu verlieren. Dementsprechend steigt f_a zunächst mit I_E an und wird bei einem gewissen Stromwert (z. B. 4 mA) konstant. In diesem Stromgebiet treten jedoch schon gewisse Hochstromeffekte [84] auf, die wieder zu einem Absinken von f_a bei hohen Strömen führen. Mit obigen Zahlen erhalten wir als resultierende Grenzfrequenz $f_a \approx 570$ MHz.

Als nächste Größe ist der Basiswiderstand $r_{BB'}$ zu betrachten. Wegen der einfacheren Rechteckgeometrie vereinfacht sich die in Abschnitt 6.3.3 angegebene Rechnung. Der Widerstand $r_{BB'}$ setzt sich hier nur aus zwei Teilen zusammen. Der erste ist der Teil zwischen $y = a_2$ und $y = a_1 + a_2$ (siehe Abb. 6.25)

$$r_1 = \frac{1}{\bar{\sigma}_1}\,\frac{a_1}{a_3 W_1} \qquad \text{(Gl. 6.144)}$$

wo $\overline{\sigma_1}$ die mittlere Leitfähigkeit der Schicht ist. Wird eine GAUSS-Verteilung

$$N_D(x) = N_D{}^0\, e^{-C^2x^2} \quad \text{(Gl. 6.145)}$$

angenommen mit $N_D(x=0) \equiv N_D{}^0$, und wegen $[N_D(x=W) \equiv N_D{}^C]$

$$C^2 \equiv \frac{1}{W^2} \ln \frac{N_D{}^0}{N_D{}^C} \quad \text{(Gl. 6.145 a)}$$

so ist

$$\overline{\sigma_1} = \frac{1}{W_1} \int_0^{W_1} e\, \mu_n(x)\, N_D(x)\, dx \approx \frac{e\, \mu_n{}^0\, N_D{}^0}{W_1} \int_0^{W_1} e^{-C^2x^2}\, dx$$

$$\approx \sigma^0 \frac{\sqrt{\pi}}{2} \left(\ln \frac{N_D{}^0}{N_D{}^C} \right)^{-\frac{1}{2}} \quad \text{(Gl. 6.146)}$$

Hier ist σ^0 die zu $N_D = N_D{}^0$, also der Oberflächendichte, gehörige Leitfähigkeit. Sie ist für $N_D{}^0 = 2{,}5 \cdot 10^{17}\ \text{cm}^{-3}$ etwa $\sigma^0 = 40\ \Omega^{-1}\,\text{cm}^{-1}$. Damit ergeben unsere typischen Zahlenwerte $\overline{\sigma_1} = 14\ \Omega^{-1}\text{cm}^{-1}$ und mit Gl. 6.144 $r_1 = 60\ \Omega$.

Der zweite Anteil zu $r_{BB'}$ kommt von dem Gebiet zwischen $y = 0$ und $y = a_2$. Dazu ist die eindimensionale Form von Gl. 6.41 zu benützen

$$\frac{d\,U(y)}{dy} = -\frac{I_B(y)}{a_3\, W\, \overline{\sigma_2}} \quad \text{(Gl. 6.147)}$$

Hier ist $\overline{\sigma_2}$ das Gl. 6.146 entsprechende Integral zwischen $x = W_1 - W$ und $x = W$ und mit unseren Zahlenwerten $\overline{\sigma_2} = 3{,}7\ \Omega^{-1}\,\text{cm}^{-1}$. Unter der Annahme, daß der Strom I_B gleichmäßig unter der gesamten Emitterfläche entspringt, wird $I_B(y) = (I_B \cdot y)/a_2$. Einsetzen in Gl. 6.147 und Integration von $y = a_2$ bis $y = x$ ergibt

$$U(y) = \frac{I_B\,(a_2{}^2 - y^2)}{2\, a_2\, a_3\, W\, \overline{\sigma_2}} \quad \text{(Gl. 6.148)}$$

Der Spannungsabfall über der ganzen Emitterbreite wird dann

$$U(0) = \frac{I_B\, a_2}{2\, a_3\, W\, \overline{\sigma_2}} \quad \text{(Gl. 6.149)}$$

Wird Gl. 6.148 in Gl. 6.40 eingesetzt, so erhält man den zweiten Anteil von $r_{BB'}$

$$r_2 = \frac{1}{a_2\, a_3} \int_0^{a_2} \frac{U(y)}{I_B}\, a_3\, dy = \frac{a_2}{3\, a_3\, W\, \overline{\sigma_2}} \quad \text{(Gl. 6.150)}$$

was numerisch $r_2 = 300\ \Omega$ ergibt. Damit ergäbe sich für den Gesamtwiderstand $r_{BB'} = r_1 + r_2 = 350\ \Omega$.

Wenn man nun dieses Ergebnis mit bei Hochfrequenz gemessenen Basiswiderständen vergleicht, so findet man, daß die tatsächlichen Werte viel niedriger liegen. Man findet $r_{BB'} \approx r_1$. Die Ursache für diesen Effekt ist die Stromverdrängung („pinch-out") des Emitterstromes zum Basiskontakt hin auf Grund des Spannungsabfalles in der Basisschicht. Dies ist in Abb. 6.26

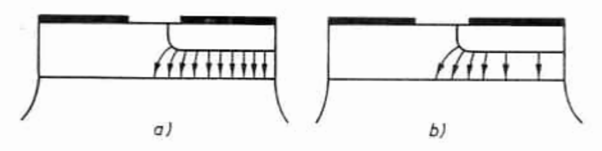


Abb. 6.26.
Zur Stromverdrängung.

schematisch dargestellt, und zwar in a) ohne Verdrängung, in b) mit Verdrängung. Bekanntlich ist ja die Emitterstromdichte an jeder Stelle y

$$j_E(y) \sim e^{\frac{U_{CB'}}{U_T}} = e^{\frac{U_{CB} - U(y)}{U_T}} \qquad \text{(Gl. 6.151)}$$

wo $U(y)$ der durch Gl. 6.148 gegebene Spannungsabfall bedeutet. Infolgedessen wird j_E für größere y [also am basisseitigen Rand, d. h. für kleinere $U(y)$], größer als für $y = 0$. Insbesondere ist das Verhältnis der Stromdichten an beiden Rändern nach Gl. 6.151 und 6.149

$$\frac{j_E(a_2)}{j_E(0)} = e^{\frac{U(0)}{U_T}} = e^{\frac{I_B\, a_2}{2 a_1 W \sigma_2 U_T}} \qquad \text{(Gl. 6.152)}$$

Der hierin vorkommende Basisstrom I_B ist $I_E\,(1 - \alpha)$, wobei die Stromverstärkung bei der Meß- bzw. Arbeitsfrequenz zu nehmen ist (also setzt sich I_B zum Teil als Verschiebungsstrom durch C_{FE} hindurch fort). Mit $\alpha = 0{,}8$ und den übrigen obigen Zahlenwerten erhalten wir z. B. schon einen Stromverdrängungsgrad $j_E(a_2)/j_E(0) = e^{3,6} = 37$. Diese Stromverdrängung zum basisnahen Rand führt zu einer wesentlichen Verkleinerung von r_2, bis bei hohen Frequenzen (kleinen α) nur noch $r_{BB'} = r_1$ übrigbleibt.

Die wirksame Kollektorkapazität mit der wirksamen Kollektorfläche $F_C \approx F_E$

$$C_{FC} \approx F_E \frac{\varepsilon\, \varepsilon_0}{l_C} \qquad \text{(Gl. 6.153)}$$

berechnet sich mit obigen Werten zu 0,12 pF. Damit wird nach Gl. 6.110a die

maximale Oszillationsfrequenz $f_{\max} = 1{,}9$ GHz. Daraus errechnet sich z. B. für eine Frequenz $f = f_a = 570$ MHz nach Gl. 6.110a eine Verstärkung $V_{L\,\mathrm{opt}} = (f_{\max}/f)^2 \approx 10$; das sind 10 dB.

Die hier beschriebene Analyse setzt voraus, daß der Kollektorbahnwiderstand klein ist. Dies ist in Wirklichkeit nur gewährleistet beim Epitaxietransistor. Andernfalls fällt die Leistungsverstärkung wegen der Notwendigkeit des Aufladens der Kollektorkapazität über den Kollektorbahnwiderstand bei höherer Frequenz stärker ab, als Gl. 6.110 angibt. Trotzdem kann die Gütezahl $f_{\max}$ auch bei Mesatransistoren, die nicht nach dem Epitaxieverfahren hergestellt sind, mit einiger Vorsicht als Dimensionierungsrichtlinie verwendet werden.

Es soll nun noch betrachtet werden, wie sich die Änderung einiger Konstruktionsparameter auf die Gütezahl $f_{\max}$ auswirkt. Zu diesem Zweck schreiben wir einmal den gesamten Ausdruck für $f_{\max}$ an, indem wir Gl. 6.143, 6.131, 6.144 und 6.153 in Gl. 6.110a einsetzen.

$$f_{\max} = \sqrt{\frac{f_a}{8\,\pi\, r_{BB'}\, C_{FC}}} \qquad \text{(Gl. 6.154)}$$

$$= \left\{ \left[\frac{\pi W^2}{D_p} \left(\frac{1}{2} \ln \frac{N_D{}^E}{N_D{}^C} \right)^{-3/2} + \frac{\pi\, l_C}{v_{\lim}} + \frac{2\,\pi\, U_T}{I_E/F_E} \right] \sqrt{\frac{\varepsilon\,\varepsilon_0\, e\, N_D{}^E}{2\,(V_D - U_{EB'})}} \right.$$

$$\left. \cdot\, 8\pi \frac{a_1}{a_3\, W_1\, \overline{\sigma_1}} \cdot \frac{a_2\, a_3\, \varepsilon\, \varepsilon_0}{l_C} \right\}^{-1/2}$$

Betrachten wir zunächst die Länge des Emitter- und Basisfleckens a_3. Diese Größe hebt sich im letzten Teil der Gleichung, d. h. in $r_{BB'} C_{FC}$ weg und tritt sonst nur in der Form $I_E/F_E = I_E/(a_2 a_3)$, also als Emitterstromdichte auf. Daraus geht hervor, daß die Verlängerung der Emitter- und Basisflecken die Möglichkeit gibt, den Emitterstrom und damit die Leistung bei gleichbleibender Stromdichte zu steigern, ohne daß in erster Näherung die Hochfrequenzleistungsverstärkung beeinträchtigt wird. Dies führt dann bei der Konstruktion von Hochfrequenzleistungsverstärkern zu einer Vielfachstreifen- oder Kamm-Struktur, wie sie in Abb. 6.27 für einen Mesatransistor gezeigt wird.

Die Gesamteindringtiefe der Basisschicht W_1 erscheint nur im Nenner des Basiswiderstandes. Man beachte, daß $\overline{\sigma_1}$ von W_1 unabhängig ist (Gl. 6.146). Infolgedessen haben wir $f_{\max} \sim \sqrt{W_1}$. Es muß also W_1 groß gemacht werden. Das geht aber nur aber im Rahmen der Forderung, daß der Abstand zwischen Emitter- und Kollektor-Übergang W klein bleiben muß, wodurch aus Fabrikationsgründen auch W_1 nicht zu groß werden kann. Im übrigen braucht W nur so klein gemacht zu werden, daß das erste Glied in der eckigen Klammer nicht merklich größer als die beiden anderen Glieder wird. Sind aber diese letzteren Glieder bestimmend, so bringt eine weitere Verkleinerung von W

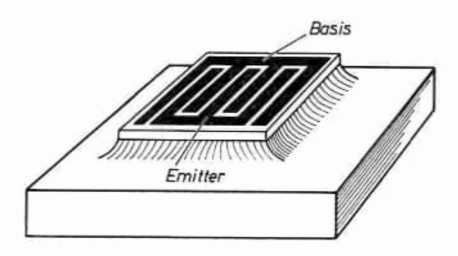


Abb. 6.27.
Kamm-Struktur eines Hochfrequenzleistungstransistors.

keinen Vorteil. In einem solchen Fall bringt möglicherweise die Absenkung von $N_D{}^E$ und damit der Emitterkapazität mehr (solange dadurch der Basiswiderstand über $\overline{\sigma_1}$ nicht verschlechtert wird).

Als nächster Parameter sei die Dicke der Kollektorsperrschicht l_C betrachtet, die nach Gl. 5.30 $\sim \sqrt{U_{CB'}/N_A{}^C}$ ist. Solange die Laufzeit durch diese Sperrschicht klein, d. h. solange das Mittelglied der eckigen Klammer klein gegenüber den anderen Gliedern bleibt, ist f_α unabhängig von l_C. Damit wird $f_{\max} \sim \sqrt{l_C}$. Bei hinreichender Vergrößerung von l_C (z. B. durch Verkleinerung der Kollektordotierung $N_A{}^C$) wird schließlich das Mittelglied bestimmend. Damit wird aber $f_{\max}$ unabhängig von l_C. Mit anderen Worten: auch hier ist der Verbesserung von $f_{\max}$ durch Vergrößern von l_C eine Grenze gesetzt.

Schließlich bleiben als wichtigste Dimensionen die Streifenbreite a_2 und der Streifenabstand a_1 übrig. Wir sehen aus Gl. 6.154, daß bei fester Emitterstromdichte der Gütefaktor $f_{\max} \sim (a_1\, a_2)^{-1/2}$ wird, d. h. je schmäler die Emitterstreifen und je kleiner ihr Abstand zum Basisstreifen ist, desto höher ist die Hochfrequenzverstärkung. Mit anderen Worten: der Weg zu besseren Höchstfrequenztransistoren muß zu immer schärferen und feineren Aufdampf- und Diffusionsmasken und zu feinerer Justierung dieser Masken führen. Mit kleiner werdender Streifenbreite wird auch das Anbringen von Kontakten an diese Streifen schwieriger. So erreicht man zwischen 10 und 50 μ eine Grenze, die mit gegenwärtig bekannten Kontaktierungsmethoden nicht unterschritten werden kann.

6.6 Der Transistor als Leistungsverstärker

Unter der Bezeichnung Leistungsverstärker faßt man alle Verstärkerelemente zusammen, bei denen außer einem gewissen Verstärkungsfaktor eine hohe Ausgangsleistung gefordert wird, die unter Umständen außerdem möglichst verzerrungsfrei sein soll. Typische Anwendungen sind Endstufen von Empfängern und Sendern sowie Steuer- und Regelstufen. Im Gegensatz zu dem in Abschnitt 6.5 Gesagten muß der Transistor in einer solchen Funktion

stark nach Spannung und Strom ausgesteuert werden, wodurch besondere Effekte auftreten, die in der bisher behandelten Theorie für kleine Amplituden nicht enthalten sind. Diese Effekte begrenzen den Aussteuerbereich nach Strom und Spannung und damit die Ausgangsleistung. Sie setzen unter Umständen die Verstärkung herab oder führen zu erhöhten Verlusten und damit zur Aufheizung des Transistors, die den Transistor oder die Schaltung zerstören könnte. Auch innerhalb dieser Grenzen wird jedoch ein gewisser Teil der elektrischen Leistung, die Verlustleistung, in Wärme umgesetzt. Diese Wärme muß durch das Gehäuse an die Umgebung abgeführt werden, um die Temperaturerhöhung im Inneren des Transistors in zulässigen Grenzen zu halten.

Wir wollen im folgenden diese an den Transistor gestellten Forderungen etwas präzisieren, wobei wir uns auf das Beispiel des A-Betriebes mit Widerstandslast in Emitterschaltung beschränken. Daran anschließend sollen die in Frage kommenden Effekte und Probleme der Reihe nach behandelt werden.

6.6.1 Grundsätzliches. Abb. 6.28 zeigt das Kennlinienfeld eines typischen Germanium-Leistungstransistors für die Emitterschaltung. Der gestrichelte Kurvenzug setzt sich zusammen aus der Hyperbel der maximal zulässigen Verlustleistung

$$I_C U_{CE} = N_{\text{verl}} \qquad \text{(Gl. 6.155)}$$

und den Geraden der maximal zulässigen Spannung

$$U_{CE} = U_{\max} \qquad \text{(Gl. 6.156)}$$

und des maximal zulässigen Kollektorstromes

$$I_C = I_{\max} \qquad \text{(Gl. 6.157)}$$

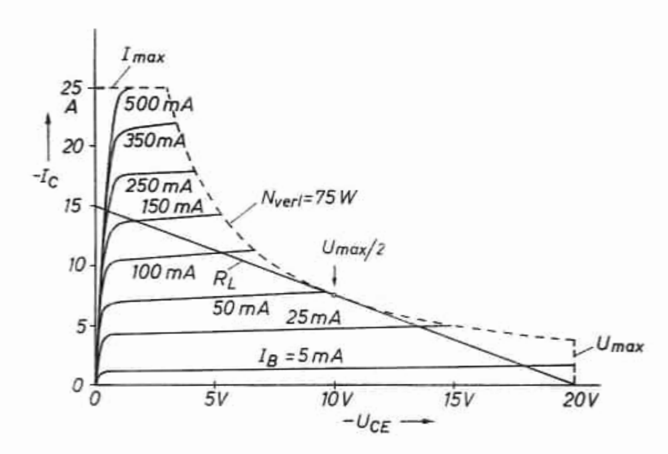


Abb. 6.28.
U_{CE}-I_C-Kennlinienfeld eines Leistungstransistors (Type CTP 1552).

Damit ist das stationär aussteuerbare Kennliniengebiet begrenzt. Wir wollen nun untersuchen, wie dadurch die Größe der Wechselstromausgangsleistung N_W begrenzt wird.

Bezeichnen wir wie bisher mit i_C und u_{CE} die Amplituden (also nicht die Effektivwerte) von Wechselstrom und -spannung, so ist bei reellem Lastwiderstand R_L die Wechselstromausgangsleistung

$$N_W = \frac{i_C\, u_{CE}}{2} = \frac{u_{CE}^2}{2\, R_L} \qquad \text{(Gl. 6.158)}$$

Um die maximal zulässige Kollektorspannung $U_{\max}$ nicht zu überschreiten, muß $u_{CE} < {}^1/_2\, U_{\max}$ sein, wobei der Gleichstromarbeitspunkt bei $U_{CE} = U_{\max}/2$ zu liegen hat. Dann wird

$$N_W \leq \frac{U_{\max}^2}{8\, R_L} \qquad \text{(Gl. 6.159)}$$

Um diese Größe bei vorgegebenem $U_{\max}$ möglichst groß zu machen, muß R_L so klein als möglich sein. Dies ist jedoch in dreierlei Hinsicht begrenzt.

Erstens muß die R_L entsprechende Arbeitsgerade ganz im erlaubten Kennliniengebiet liegen; d. h. sie darf, von Punkt ($U_{CE} = U_{\max}$, $I_C = 0$) ausgehend, die Verlustleistungshyperbel höchstens berühren, wie das in Abb. 6.28 gezeichnet ist. Der Berührungspunkt berechnet sich aus Gl. 6.155 und Abb. 6.28 in elementarer Weise zu $U_{CE} = U_{\max}/2$. Dann ist R_L begrenzt zu

$$R_L \geq \frac{U_{\max}^2}{4\, N_{\text{verl}}} \qquad \text{(Gl. 6.160)}$$

was, eingesetzt in Gl. 6.159, eine Grenze für N_W ergibt.

$$N_W \lesssim \frac{N_{\text{verl}}}{2} \qquad \text{(Gl. 6.161)}$$

Damit haben wir auch gleich den maximalen Wirkungsgrad des A-Betriebes berechnet, nämlich

$$N_W/N_{\text{Gleichstrom}} = N_W/N_{\text{verl}} \leq 50\%$$

Eine zweite Begrenzung erfährt der Lastwiderstand R_L nach unten durch den Maximalstrom; d. h. der Ordinatenabschnitt der Geraden R_L muß $\leq I_{\max}$ sein, also $U_{\max}/R_L \leq I_{\max}$. Das ergibt mit Gl. 6.159 die Bedingung

$$N_W \leq \frac{U_{\max}\, I_{\max}}{8} \qquad \text{(Gl. 6.162)}$$

Drittens wird R_L durch die geforderte Leistungsverstärkung nach unten begrenzt. Eine solche Forderung spielt bei NF-Anwendungen allerdings eine

untergeordnete Rolle. Wir verwenden Gl. 6.100 mit den Indizes E bei den y-Koeffizienten (Emitterschaltung). Dabei kann im allgemeinen bei Leistungsverstärkern die Näherung $1/R_L = y_L \gg y_{22E}, y_{12E}$ und $y'_{11} \gg 1/r_{BB'}$ benutzt werden, also erhebliche Fehlanpassung und hohe Stromdichte. Man erhält dann die Leistungsverstärkung

$$\overline{V}_L = \left| \frac{y_{21E}}{1/R_L} \right|^2 \cdot \frac{1/R_L}{\Re e\,(y_{11E})} = \left| \frac{y_{21E}}{y_{11E}} \right|^2 \cdot \frac{R_L}{\Re e\,(1/y_{11E})}$$

Unter Benützung von Gl. 6.119, 6.115 und 6.74 bis 6.79 erhält man

$$\overline{V}_L \approx \beta^2 \frac{R_L}{r_{BB'}} \qquad \text{(Gl. 6.163)}$$

R_L darf also nicht kleiner gemacht werden, als die geforderte Leistungsverstärkung $\overline{V}_L$ erlaubt. Durch Kombination mit Gl. 6.159 ergibt sich die dritte Begrenzung für die Ausgangsleistung

$$N_W \leq \frac{\beta^2\, U^2{}_{\max}}{8\, \overline{V}_L\, r_{BB'}} \qquad \text{(Gl. 6.164)}$$

Die Gleichungen 6.161, 6.162 und für HF-Endstufen Gl. 6.164 stellen demnach drei Bedingungen für N_W dar, deren jede einzelne N_W begrenzen kann. Dementsprechend muß für einen guten Leistungstransistor N_{verl}, $U_{\max}$, $I_{\max}$ und β groß, $r_{BB'}$ dagegen klein sein, was auch ohne Rechnung plausibel ist.

6.6.2 Verlustleistung und Wärmeabfuhr. In Abschnitt 6.2.9 wurde schon erklärt, daß ein stabiler Betrieb von Transistoren oberhalb einer bestimmten Sperrschichttemperatur nicht mehr möglich ist. Da jedoch durch die Verlustleistung am Kollektor ständig Wärme entsteht, ist es notwendig, diese so gut wie möglich an die Umgebung abzuführen. In Analogie mit dem elektrischen Stromfluß definiert man für den Wärmefluß einen Wärmewiderstand K durch die Gleichung:

$$N_{\text{verl}} = \frac{T_i - T_a}{K} \qquad \text{(Gl. 6.165)}$$

wobei T_i die innere Temperatur am Kollektorübergang und T_a die äußere, die Umgebungstemperatur ist. Der Wärmewiderstand K wird in °C/Watt gemessen und kann als Summe aller Teilwiderstände aufgefaßt werden, z. B. der Wärmewiderstände zwischen Kollektorübergang und Innenseite des Gehäusebodens, zwischen Gehäuseboden innen und außen, zwischen Gehäuse außen und der Chassisplatte oder der Kühlrippe, auf der der Transistor festgeschraubt wird. Die maximale Kollektortemperatur T_i liegt fest: bei Ge etwa 100° C, bei Si wegen des an sich kleinen Sättigungsstromes höher, etwa 250° C. Ebenso liegt T_a fest; und da wegen Gl. 6.161 eine möglichst hohe zulässige Verlustleistung N_{verl} gefordert wird, muß K so klein wie möglich sein.

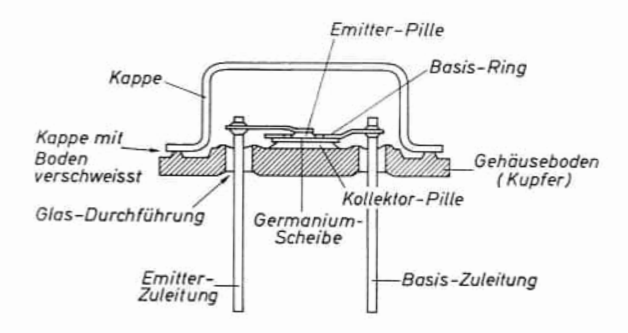


Abb. 6.29.
Legierter Germanium-Leistungstransistor
vom Typ 2N 1146.

In Abb. 6.29 ist der Aufbau eines typischen legierten Germanium-Leistungstransistors dargestellt. Wie man sieht, ist die Kollektorlegierungspille direkt auf den Kupferboden des Gehäuses aufgelötet. Die Legierungspille selbst ist so dünn gehalten, wie technologisch möglich, oder sie ist sogar zwischen Legieren und Auflöten zum Teil abgeschnitten worden, da sonst das Legierungsmetall (meist Indium oder Blei) einen zu hohen Wärmewiderstand ergeben würde. Das Gehäuse selbst wird mit seinem flachen Kupferboden durch kräftige Schrauben am Chassisblech oder an einem besonderen Kühlkörper festgeschraubt.

Bezüglich des Wärmewiderstandes sind auch die diffundierten Transistoren besonders gut geeignet, da sich hier der Kollektorübergang sowieso nahe an der Grundplatte befindet, sofern das Halbleiterplättchen nicht zu dick ist (vgl. Abb. 6.24 und 6.27).

In praktischen Leistungstransistoren erreicht man im allgemeinen $K = 0{,}8\ ^\circ\mathrm{C/W}$. Das ergibt nach Gl. 6.165 z. B. bei Germanium etwa 100 Watt, bei Silizium etwa 290 Watt. Man erreicht also mit Silizium wesentlich höhere maximale Verlustleistungen als mit Germanium.

Das in Gl. 6.161 vorkommende N_{verl} umfaßt natürlich nur die unvermeidliche Verlustleistung. Alle Serienwiderstände, z. B. im Kollektor, Emitter oder Basis ($r_{BB'}$), ebenso wie ein erhöhter Kollektorstrom (Leckstrom oder beim Spannungsdurchbruch, siehe unten!) erhöhen lediglich die Erwärmung. Sie führen deshalb nur zur Verkleinerung der „nutzbringenden“ Verlustleistung in Gl. 6.161 und müssen deshalb so klein wie möglich gehalten werden.

6.6.3 Spannungsdurchbruch. Wir wenden uns jetzt der Frage zu, welche Effekte die maximal zulässige Kollektorspannung $U_{\max}$ bestimmen. Bei einer Steigerung der Kollektorspannung beobachtet man bei einem gewissen Punkt ein sehr steiles Ansteigen des Kollektorstromes, das nach den obigen Ausführungen zu

einer starken Überhitzung des Transistors führen kann und deshalb vermieden werden muß. Dieser Anstieg kann verschiedene Ursachen haben.

Die erste ist der in Abschnitt 5.7.2 erwähnte Lawinen-(oder Avalanche-) Durchbruch. Dabei werden bei einem hinreichend großen Feld F_{av} in der Sperrschicht die Träger so viel Energie aufnehmen, daß sie durch Stoß weitere Trägerpaare erzeugen, was zum lawinenartigen Anstieg des Stromes führt [74, 75]. Wir wollen das Problem hier nur in sehr qualitativer Weise andeuten.

Man kann annehmen, daß dieser Lawinendurchbruch bei einer ganz bestimmten kritischen Feldstärke F_{av} in der Sperrschicht erfolgt. Die in der Sperrschicht auftretende maximale Spannung ist für den abrupten Übergang durch Gl. 5.31 und Gl. 5.30 gegeben. Also ist beim Durchbruch für $U = -U_{av} \gg V_D$ und $N_A \gg N_D$

$$F_{av} = F_{\max} \sim \sqrt{N_D\, U_{av}}$$

Demnach wäre die „Avalanche"-Spannung U_{av} proportional $1/N_D$. Empirisch findet man für abrupte Übergänge in Germanium

$$U_{av} \approx 1{,}5 \cdot 10^{14}\, N_D^{-0{,}8}\ [\text{Volt}] \qquad \text{(Gl. 6.166)}$$

wobei N_D in cm^{-3} gemessen wird.

Wir haben bisher so getan, als ob der Avalanche-Durchbruch ausschließlich in homogenem Material stattfinden würde. Tatsächlich tritt der Durchbruch bevorzugt an schwachen Stellen des PN-Überganges auf, d. h. dort, wo $F_{\max}$ besonders hoch ist. Das ist insbesondere an der Oberfläche der Fall. Man muß also durch sorgfältiges Ätzen, Trocknen und Verkapseln versuchen, diesen Effekt so gut wie möglich zu beherrschen. Der Oberflächendurchbruch tritt aus noch nicht ganz geklärten Gründen bei Silizium erst bei höheren Spannungen auf als bei Germanium, ein weiterer Grund, weshalb Silizium in neuerer Zeit für Leistungstransistoren bevorzugt wird.

Es kann aber noch ein anderer Effekt auftreten, der zu einem Durchbruch bei hoher Spannung führt und somit $U_{\max}$ bestimmen kann, nämlich die sogenannte Sperrschichtberührung, oder „Punch-through". Bei zunehmender Kollektorspannung dehnt sich bekanntlich nach Gl. 5.30 die Kollektorsperrschicht immer mehr aus und berührt schließlich bei der „Punch-through"-Spannung U_{pt} die Emitter-Raumladungsschicht. Eine weitere Erhöhung der Kollektorspannung führt dann zu einer sehr raschen Erniedrigung der Potentialschwelle des Emitterübergangs und damit zu einem starken Anstieg eines Stromes von Löchern aus dem Emitter, die dann direkt in die Raumladungszone des Kollektors hineingezogen werden. Im Falle eines legierten PNP-Transistors haben wir einen unsymmetrischen, abrupten Kollektorübergang mit $N_A \gg N_D$. Dann ist nach Gl. 5.30

$$l_C = \sqrt{\frac{2\,\varepsilon\,\varepsilon_0}{e\,N_D}\,(-U)} \qquad \text{(Gl. 6.167)}$$

Wenn $U = -U_{pt}$, ist $l_C \approx W$, und wir erhalten

$$U_{pt} = \frac{e}{2\,\varepsilon\,\varepsilon_0}\, W^2 N_D \qquad \text{(Gl. 6.168)}$$

Man sieht, daß für diesen Effekt die Abbruchspannung sich mit der Basisdotierung gerade umgekehrt ändert wie beim Lawineneffekt. Mit anderen Worten, bei sehr hochohmiger Basis tritt zuerst „Punch-through", bei sehr niederohmiger Basis zuerst Avalanche auf. Die tatsächliche Wahl der Basisdotierung ist also wieder ein Kompromiß, in den aber selbstverständlich die Forderung nach einem kleinen Basiswiderstand $r_{BB'}$ (siehe Gl. 6.164) eingeschlossen werden muß. Weiterhin ist bezüglich der Wahl von W ein Kompromiß zwischen U_{pt}, α (oder β) und f_α zu schließen, wie schon in Abschnitt 6.5.2 ausgeführt wurde.

Beim diffundierten Transistor werden die Verhältnisse etwas komplizierter. Dann folgt die Avalanche-Spannung einem anderen Gesetz als in Gl. 6.166 angedeutet, und es geht unter anderem der Dotierungsgradient am Kollektorübergang ein. Außerdem ist die Ausbreitung der Kollektorsperrschicht gegen den Emitter hin wegen der in dieser Richtung ansteigenden Basisdotierung begrenzt. Sie dehnt sich vielmehr bevorzugt in die Kollektorzone hinein aus. Demzufolge ist bei diesen Transistoren die Avalanche-Spannung in der Regel allein ausschlaggebend, da der „Punch-through" weitgehend unmöglich gemacht ist.

6.6.4 Effekte bei hohen Stromdichten

6.6.4.1 Die Stromabhängigkeit von α: In Abschnitt 6.3.4, Gl. 6.46 bis Gl. 6.48, wurde schon ein Ausdruck für die Stromverstärkung in einem dreidimensionalen Modell abgeleitet. Wegen der Näherung Gl. 6.46 gilt Gl. 6.48 auch für die Wechselstromverstärkung bei tiefen Frequenzen, d. h. wir können

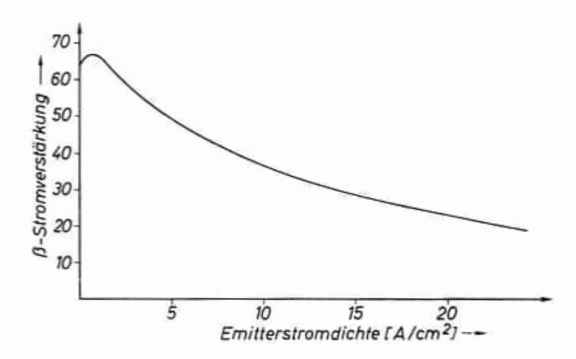


Abb. 6.30.
β-Stromverstärkung in Abhängigkeit von der Emitterstromdichte.

darin A_N durch α ersetzen. Wir sehen daraus, daß nach der bisherigen Theorie die Stromverstärkung vom Strom unabhängig sein sollte. Dem widerspricht die experimentelle Erfahrung, wie in Abb. 6.30 dargestellt ist. Demnach steigt α (und damit auch β) bei kleinen Strömen etwas an, geht durch ein Maximum und beginnt dann kontinuierlich zu fallen. Der Grund für das Versagen der einfachen Theorie in diesem Punkt liegt daran, daß dabei stets vorausgesetzt wurde, daß die injizierten Minoritätendichten in der Basis klein gegenüber den Majoritätendichten sind, und daß der Basisstrom vernachlässigbar klein ist. Das ist offenbar bei höheren Stromdichten nicht mehr der Fall, und wir haben Gl. 6.48 zu korrigieren. Wir folgen dabei weitgehend der Arbeit von WEBSTER [83].

Zunächst haben wir zu berücksichtigen, daß infolge der Löcherinjektion vom Emitter her am Emitter eine höhere Löcherdichte herrscht als am Kollektor. Wegen der Forderung der Raumladungsneutralität in der Basis ($n = N_D + p$) stellt sich auch eine höhere Elektronendichte am Emitter ein. Wenn p nicht mehr klein ist, wird auch die Inhomogenität von n nicht mehr zu vernachlässigen sein. Diese hat wegen ihrer Diffusionstendenz entsprechend dem beim Drifttransistor Gesagten (Abschnitt 6.5.3) nach Gl. 6.136 ein Feld zur Folge

$$-\nabla\varphi = -U_T \frac{1}{n} \nabla n$$

und wegen der Ladungsneutralität

$$-\nabla\varphi = -U_T \frac{1}{N_D + p} \nabla p \qquad \text{(Gl. 6.169)}$$

Setzt man dies in die Stromgleichung der Löcherstromdichte (Gl. 4.50) ein, so erhält man

$$\boldsymbol{j}_p = -e \frac{p}{N_D + p} \mu_p U_T \nabla p - e D_p \nabla p$$

$$= -e \left(\frac{p}{N_D + p} + 1 \right) D_p \nabla p \qquad \text{(Gl. 6.170)}$$

Dieselbe Gleichung kann auch aus Gl. 5.41 mit 5.42 für $\boldsymbol{j}_n = 0$, also $\boldsymbol{j} = \boldsymbol{j}_p$, abgeleitet werden.

Wie man sieht, hat das Elektronendriftfeld eine Vergrößerung der effektiven Diffusionskonstanten der Löcher um dem Faktor $p/(N_D + p) + 1$ zur Folge. Für sehr starke Injektionen, d. h. für $p \gg N_D$, geht dieser gegen 2. Die Integration von Gl. 6.170 von $x = 0$ bis $x = W$ und von $p = p(0)$ bis $p = 0$ ergibt j_p als Funktion der Löcherdichte $p(0)$ am emitterseitigen Rand der Basis. Damit wird der Emitterstrom

$$\frac{I_E}{F_E} = j_p = \frac{e D_p}{W} p(0) \left[2 - \frac{N_D}{p(0)} \ln \left(\frac{p(0)}{N_D} + 1 \right) \right] \qquad \text{(Gl. 6.171)}$$

Diese Gleichung hat nun bei der Ableitung der allgemeineren Gl. 6.48 an die Stelle von Gl. 6.8a zu treten.

Weiterhin haben wir bei der Gl. 6.47 im zweiten Glied, dem Elektronenstrom I_{En}, zu berücksichtigen, daß auch die Elektronenkonzentration im Emitter durch die Vergrößerung von $n_N = N_D$ nach $N_D + p$ angehoben wird. Nach Gl. 5.3 ist $n_{PE} \sim n_N$ und geht deshalb für $n_N = N_D + p$ über in

$$n_{PE} \to n_{PE} \frac{N_D + p(0)}{N_D} \qquad \text{(Gl. 6.172)}$$

Außerdem ist zu beachten, daß $e^{U_{EB'}/U_T}$ mit I_E und damit $p(0)$ verknüpft ist durch

$$e^{U_{EB'}/U_T} = \frac{p(0)}{p_N} \qquad \text{(Gl. 6.173)}$$

Das Volumenrekombinationsglied wollen wir in der hier gewählten Näherung vernachlässigen.

Dann wird unter Berücksichtigung der Gleichungen 6.171 bis 6.173 aus Gl. 6.47

$$\alpha = \frac{\partial I_C}{\partial I_E} \approx 1 - \frac{\partial I_{En}}{\partial I_E} - \frac{\partial I_{pO}}{\partial I_E}$$

$$= 1 - \frac{D_n n_{PE} W}{D_p p_N L_{nE}} \frac{\partial\left\{\left(1 + \frac{p(0)}{N_D}\right) p(0)\right\}}{\partial\left\{p(0)\left[2 - \frac{N_D}{p(0)} \ln\left(1 + \frac{p(0)}{N_D}\right)\right]\right\}}$$

$$- \frac{2b^* W s}{a D_p} \frac{\partial p(0)}{\partial\left\{p(0)\left[2 - \frac{N_D}{p(0)} \ln\left(1 + \frac{p(0)}{N_D}\right)\right]\right\}}$$

$$= 1 - \frac{\sigma_N W}{\sigma_{PE} L_{nE}}\left(1 + \frac{p(0)}{N_D}\right) - \frac{2\mathrm{b}^* W s}{a D_p} \frac{1 + \frac{p(0)}{N_D}}{1 + 2\frac{p(0)}{N_D}} \qquad \text{(Gl. 6.174)}$$

Dieses Ergebnis kann in nicht ganz strenger, aber wesentlich einfacherer Weise auch dadurch gewonnen werden, daß man in Gl. 6.48 die Größe σ_N ($\sim n_N$) mit dem Faktor $[N_D + p(0)] / N_D$ von Gl. 6.172 und die Größe D_p mit dem Faktor $p(0) / [(N_D + p(0)] + 1$ von Gl. 6.170 versieht.

Diese Faktoren (und damit die Änderung der Stromverstärkung) werden wesentlich, wenn $p(0)$ in die Größenordnung von N_D kommt. Für die in

Abb. 6.8 gezeichneten Gleichgewichtsdichten ist $p(0) = N_D$ bei $U_{EB'} = 0{,}23$ V oder (für $W = 40\,\mu$) bei 1,8 A/cm² (vgl. Abb. 6.30).

Wie man sieht, nimmt das letzte Glied von Gl. 6.174 mit $p(0)$, also auch mit I_E, ab. Folglich nimmt α mit I_E zunächst zu. Der Grund ist leicht aus der obigen Ableitung zu sehen: Durch das mit I_E zunehmende Driftfeld werden die Löcher stärker zum Kollektor gezogen und so daran gehindert, zur Oberfläche zu gelangen. Die Oberflächenrekombination nimmt ab. Das erklärt den ersten Anstieg bis zum Maximum. Dann überwiegt die Zunahme des zweiten Gliedes, das die Emitterergiebigkeit absenkt infolge der Erhöhung der Elektronenkonzentration in der Basis. Das führt zur entscheidenden Abnahme von α und β. Der genaue Zusammenhang zwischen $p(0)$ und I_E ist durch Gl. 6.171 gegeben.

Wir wollen dies jedoch hier nicht explizit weiterverfolgen, sondern lediglich überlegen, welche Maßnahmen den störenden α-Abfall verringern und damit $I_{\max}$ vergrößern können. Es muß dazu offenbar das zweite Glied von Gl. 6.174 klein gemacht werden. Es sollte also erstens die Basisdicke W klein sein. Diese Größe ist aber in der Regel schon durch den oben erwähnten Kompromiß zwischen β und $U_{\max}$ sowie durch die Technologie so festgelegt, daß nicht mehr viel zu variieren ist. Ebenfalls festgelegt durch den technologischen Prozeß ist $L_{nE} = D_n \tau_{nE}$ in der Emitterzone. So bleibt nur der Weg, σ_{PE}/σ_N so groß wie möglich zu machen, also eine Vergrößerung von γ. Dies geschieht beim diffundierten Silizium-Transistor durch eine sehr steile und wenig tiefe Diffusion des Emitters. Beim legierten Germanium-Transistor wird die Leitfähigkeit σ_{PE} der Emitterzone desto größer sein, je größer der Segregationskoeffizient des dotierenden Stoffes in der Emitterpille ist, d. h. je größer die Konzentration ist, mit der er in die Rekristallisationsschicht eingebaut wird. Indium, das übliche Pillenmaterial für *PNP*-Germanium-Transistoren, hat z. B. einen kleinen

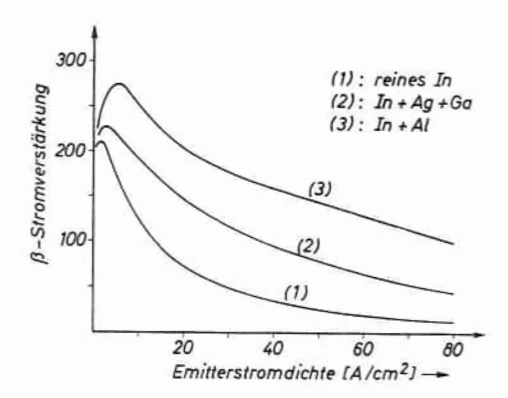


Abb. 6.31.
β-Stromverstärkung in Abhängigkeit von der Emitterstromdichte für verschiedene Legierungssubstanzen [92].

Segregationskoeffizienten. Deshalb fügt man gewöhnlich Gallium oder Aluminium oder beides zu [92], die beide in dieser Hinsicht etwa 100fach besser als Indium allein sind. Dieser Effekt ist in Abb. 6.31 klar zu sehen.

Diese WEBSTERsche Theorie wurde in neuerer Zeit ergänzt durch WANG und WU [89], die insbesondere den Einfluß der Stromverdrängung (pinch-out), die in Abschnitt 6.5.4.5 schon erwähnt wurde, berücksichtigen. Wie dort gezeigt wurde, erzeugt der hohe Basisstrom längs der Basisschicht einen Spannungsabfall, der dazu führt, daß nur die Randgebiete der Emitterfläche injizieren. Während beim Hochfrequenztransistor der hohe Basisstrom vom Verschiebungsstrom durch die Emittersperrschichtkapazität herrührt, liegt hier allgemein wegen des hohen Stromes ein hoher Basisstrom vor. Dieser Effekt nimmt natürlich mit dem Strom zu und begünstigt die Oberflächenrekombination, wie man z. B. an Hand von Abb. 6.26 b leicht einsehen kann.

Einen weiteren Beitrag zum Abfall der Stromverstärkung bei hohem Strom liefert die Tatsache, daß bei Stromverdrängung (Abb. 6.26 b) die Strompfade der Löcher länger und damit die Volumenrekombination größer wird.

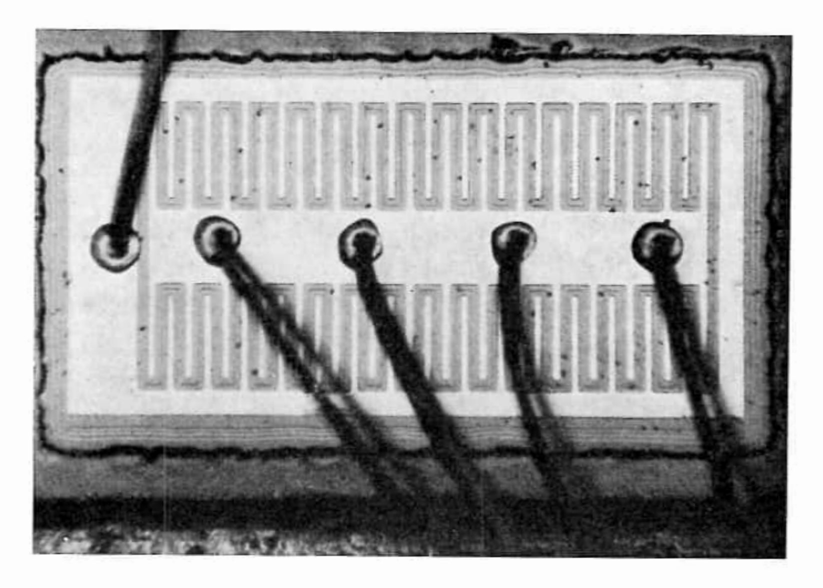
Es ist leicht einzusehen, daß bei gegebenem Emitterstrom der Basisstrom und damit die beiden letzten Effekte desto kleiner sind, je größer γ, also σ_{PE}/σ_N, ist. D. h. der Erfolg der oben beschriebenen technologischen Maßnahmen kann auch zur Bestätigung der Theorie von WANG und WU herangezogen werden.

Bei Drifttransistoren führt schließlich die starke Injektion bei hohem Strom zu einer Verwaschung (Überschwemmung) des von ∇N_D stammenden Driftfeldes und führt dort auch zu einer Reduktion von α (und außerdem von f_α) bei hohen Strömen.

6.6.4.2 Wirksame Emitterfläche und Basiswiderstand bei hohem Strom: Die oben beschriebene Verdrängung des Emitterstromes zum basisseitigen Rand des Emitters führt aber nicht nur zu einer Verkleinerung des Basiswiderstandes (Abschnitt 6.5.4.5) und zu einer Erhöhung der Oberflächenrekombination (Abschnitt 6.6.4.1), sondern auch zu einer Verkleinerung der wirksamen Emitterfläche [90]. Das bedeutet, daß für einen gegebenen Emitterstrom I_E und eine tatsächliche Emitterfläche F_E die Stromdichte j_p insbesondere am Rand wesentlich größer als I_E/F_E ist. Dadurch werden die oben beschriebenen Hochstromeffekte schon bei kleineren Strömen wirksam, als es nach der obigen Rechnung (Gl. 6.171) der Fall sein müßte.

Dementsprechend können die wirksamen Emitterflächen nicht wesentlich vergrößert und die Hochstromeffekte nicht wesentlich verkleinert werden, indem man die Fläche F_E schlechthin vergrößert, z. B. durch Verbreiterung des Emitterstreifens. Sondern man muß ganz analog zum Hochfrequenztransistor die Längsdimension des Emitters vergrößern, also den Emitterrand verlängern, um bei gegebenem Strom I_E die Stromdichte j_p klein zu halten. Man gelangt

wieder, wie bei den Hochfrequenztransistoren, zu Kammstrukturen (Abb. 6.27) oder (für legierte Transistoren besser geeignet) zu konzentrischen Emitter- und Basisringen. Der einfachste Fall ist der in Abb. 6.29 gezeigte runde Emitter mit Basisring. Tafel 18 zeigt die Aufsicht auf einen modernen diffundierten Silizium Leistungstransistor.



Tafel 18.
Aufsicht auf einen diffundierten Silizium-Leistungstransistor

Schließlich ist noch ein weiterer Hochstromeffekt zu erwähnen, der zu einer Verkleinerung des Basiswiderstandes führt, die Leitfähigkeitsmodulation für den Querstrom I_B in der Basis. Die Leitfähigkeit σ_N, die in den Basiswiderstand $r_{BB'}$ (z. B. Gl. 6.41, 6.43) eingeht, ist bei hoher Injektion nicht mehr gleich $e\mu_n N_D$, sondern $e\mu_n (N_D + p)$, und als solche wegen p wesentlich höher und überdies nach dem bisher Gesagten vom Ort abhängig.

Zum Abschluß sei noch darauf hingewiesen, daß im Transistor bei großer Aussteuerung eine ganze Reihe weiterer Effekte auftreten können, bezüglich derer wir jedoch nur auf die Literatur [94, 84] verweisen wollen.

Literaturverzeichnis

Lehrbücher zum weiteren Studium

E. Spenke: Elektronische Halbleiter (Springer, 1956).
W. Shockley: Electrons and Holes in Semiconductors (Van Nostrand 1953).
H. A. Müser: Einführung in die Halbleiterphysik (Steinkopff, 1960).
R. A. Smith: Semiconductors (University Press, Cambridge, 1959).
W. C. Dunlap jr.: An Introduction to Semiconductors (John Wiley, 1957).
C. Kittel: Introduction to Solid State Physics (John Wiley, 1956).
W. Ehrenberg: Electric Conduction in Semiconductors and Metals (Clarendon Press, Oxford, 1958).
J. Dosse: Der Transistor (Oldenbourg, 1957).
O. Madelung: Halbleiter. In: Flügge, Handbuch der Physik XX (Springer, 1957).
L. P. Hunter: Handbook of Semiconductor Electronics (McGraw Hill, 1962).
W. W. Gärtner: Transistors, Principles, Design and Applications (Van Nostrand, 1960).
G. Rusche, K. Wagner und F. Weitzsch: Flächentransistoren (Springer, 1961).
R. D. Middlebrook: An Introduction to Junction Transistor Theory (J. Wiley and Sons, 1957).

Literaturzitate des Textes

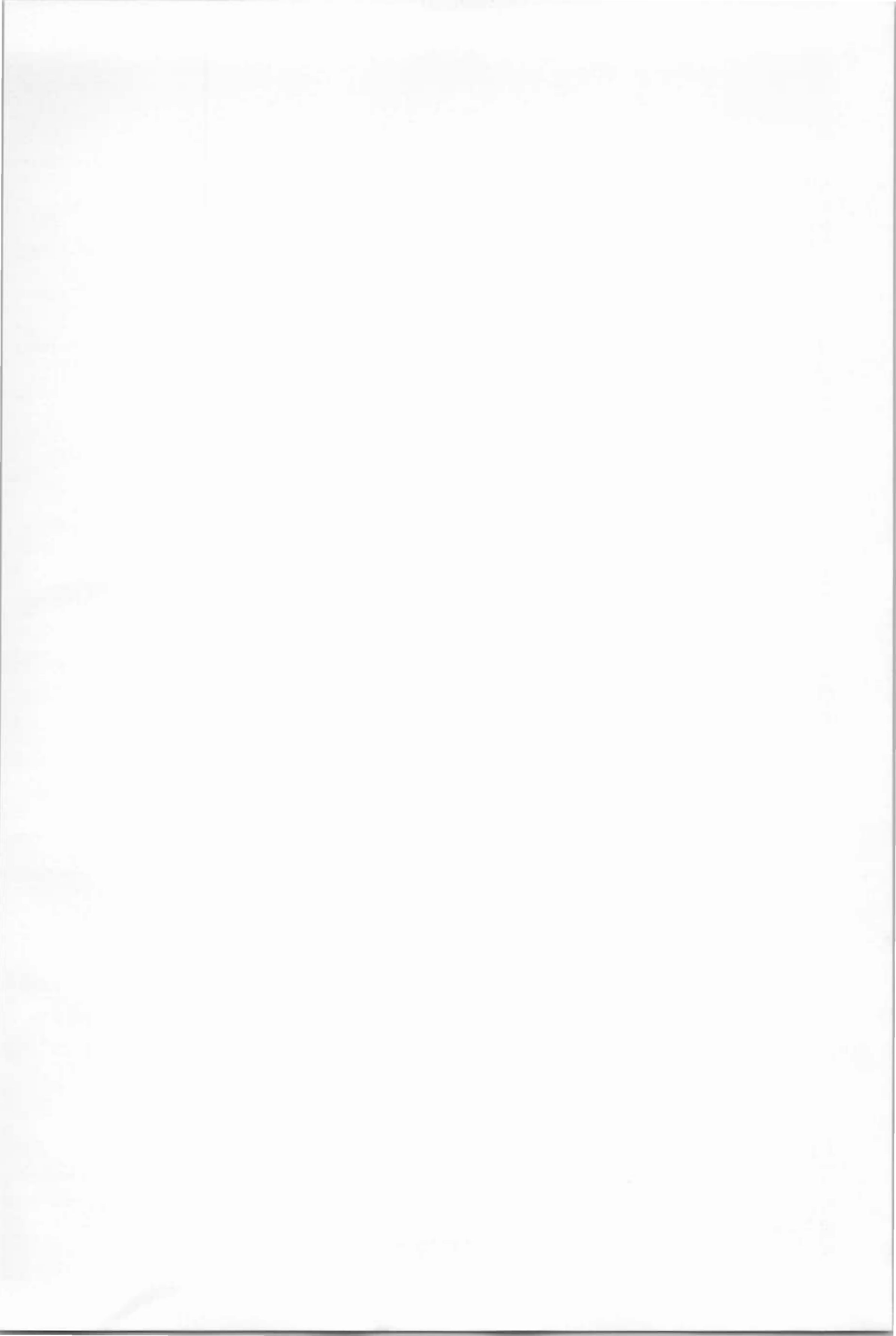
1. R. W. Pohl: Einführung in die Elektrizitätslehre (Springer, 1960); Einführung in die Optik (Springer, 1958).
2. W. G. Pfann: Zone Melting (John Wiley 1958).
3. P. H. Keck, M. J. E. Golay: Phys. Rev. **89**, 1297 (1953).
4. R. Emeis: Z. Naturforsch. **9a**, 67 (1954).
5. S. Müller: Z. Naturforsch. **9b**, 504 (1954).
6. J. Czochralski: Z. phys. Chem. **92**, 219 (1917).
7. G. K. Teal, J. B. Little: Phys. Rev. **78**, 647 (1950).
8. G. K. Teal, E. Buehler: Phys. Rev. **87**, 190 (1952).
9. H. Kleinknecht, K. Seiler: Z. Physik **139**, 599 (1954).
10. F. D. Rosi: RCA-Review **19**, 394 (1958).
11. W. C. Dash: J. Appl. Phys. **30**, 459 (1959).
12. W. C. Dash: J. Appl. Phys. **27**, 1193 (1956).
13. H. Rieger, K. Seiler: Z. Naturforsch. **16a**, 220 (1961).
14. W. C. Dunlap jr.: Phys. Rev. **96**, 40 (1954).
15. P. P. Debye, E. M. Conwell: Phys. Rev. **93**, 693 (1954).
16. W. C. Dunlap jr., D. E. McMillan, R. A. Brooks: Phys. Rev. **93**, 911 (1954).
17. O. Madelung: Ergeb. exakt. Naturwiss. **27**, 92 (1953).
18. F. J. Morin, J. P. Maita: Phys. Rev. **94**, 1525 (1954).
19. K. Schuegraf, K. Seiler: Z. Natforsch. **15a**, 368 (1960).
20. H. Kronmüller, J. Jaumann, K. Seiler: Z. Naturforsch. **11a**, 243 (1956).
21. E. M. Conwell: Proc. IRE **46**, 1281 (1958).
22. H. Brooks: Advances in Electronics and Electron Physics **7** (New York Academic Press, Inc., 1955).
23. E. S. Rittner: Phys. Rev. **101**, 1291 (1956).

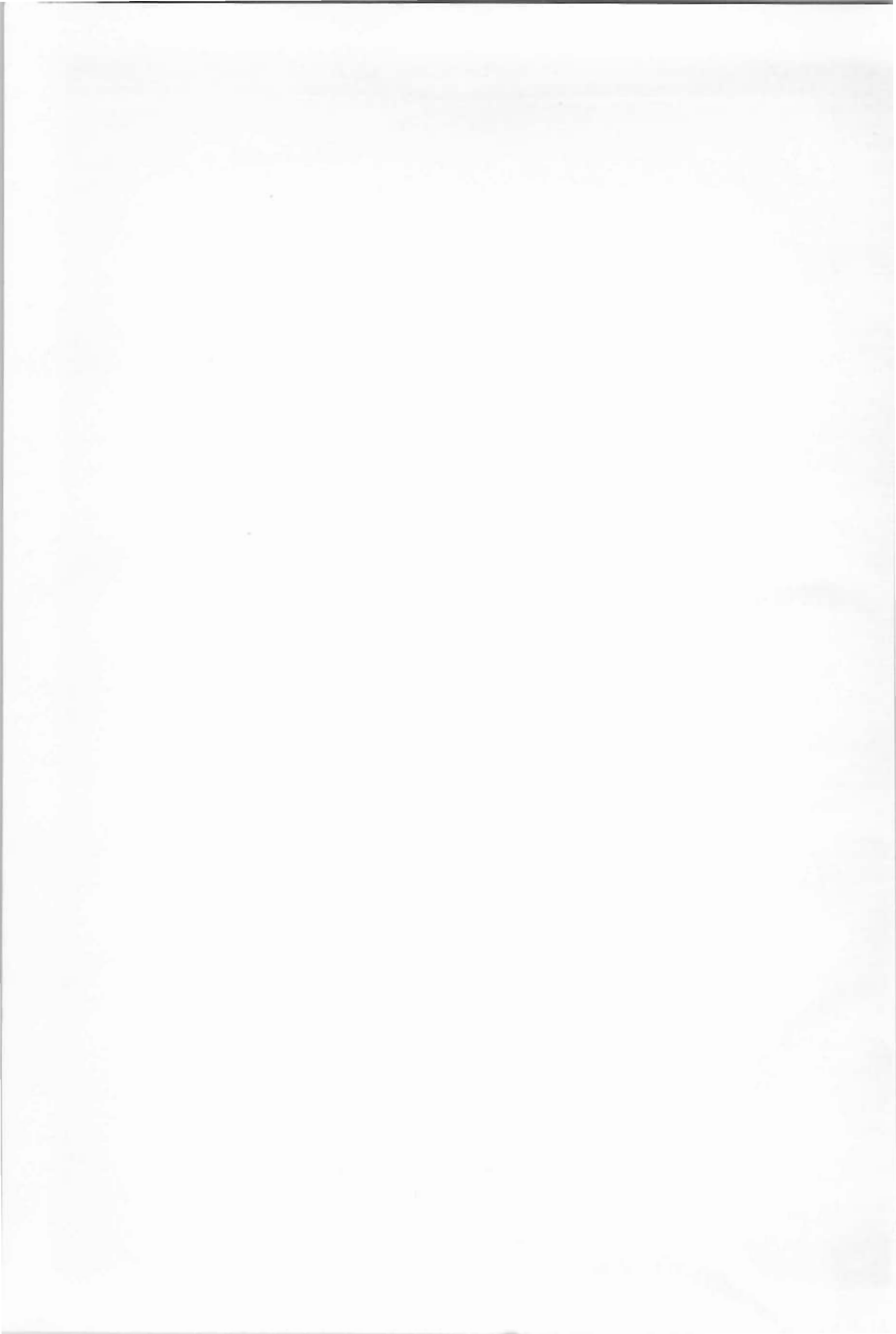
24. A. Sommerfeld, H. Bethe: In: Geiger-Scheel, Handbuch d. Phys. **XXIV, 2**
25. E. Spenke: Elektronische Halbleiter, Anhang III. (Springer 1956)
26. D. T. Stevenson, R. J. Keyes: J. Appl. Phys. **26**, 190 (1955).
27. M. B. Prince: Phys. Rev. **92**, 681 (1953).
28. R. H. Kingston: Semiconductor Surface Physics (Pensylvania University Press 1957).
29. W. Bösenberg: Z. Naturforsch. **10a**, 285 (1955).
30. C. F. Fuller, J. A. Ditzenberger: J. Appl. Phys. **27**, 544 (1956).
31. K. Hubner, W. Shockley: Phys. Rev. Lett. **4**, 504 (1960).
32. A. F. Joffe: Halbleiter-Thermoelemente (Akademie-Verlag 1957).
33. W. Shockley: Bell Syst. Tech. J. **28**, 435 (1949).
34. F. S. Goucher, G. L. Pearson, M. Sparks, G. K. Teal, W. Shockley: Phys. Rev. **81**, 637 (1951).
35. K. Seiler: A.T.M. **207, Z 52–1**, 87 (1953).
36. C. T. Sah, R. N. Noyce, W. Shockley: Proc. IRE **45**, 1228 (1957).
37. J. Bardeen, W. H. Brattain: Phys. Rev. **74**, 230 (1948).
38. J. M. Early: Bell Syst. Tech. J. **33**, 517 (1954).
39. H. Krömer: Archiv elektr. Übertragung **8**, 223 (1954), **8**, 363 (1954), **8** und 499 (1954).
40. J. L. Moll, M. Tannenbaum, J. M. Goldey, N. Holonyak: Proc. IRE **44**, 1174 (1956).
41. W. Shockley: Electronics Industries, August 1957.
42. W. Shockley, M. Sparks, G. K. Teal: Phys. Rev. **83**, 151 (1951).
43. W. Schottky: Z. Physik **118**, 539 (1942).
44. G. L. Pearson, W. T. Read, W. Shockley: Phys. Rev. **85**, 1055 (1952).
45. W. Shockley, W. T. Read: Phys. Rev. **87**, 835 (1952).
46. A. Hoffmann: Halbleiterprobleme **2**, 106 (1955).
47. K. Raithel, F. Spitzer: SEG-Nachrichten **1**, (1954).
48. C. Wagner: Phys. Z. **32**, 641 (1931).
49. D. Geist: A.T.M. **227, J 392–2**, 281 (1954).
50. D. A. Jenny, J. J. Loferski, P. Rappaport: Phys. Rev. **101**, 1208 (1956).
51. F. S. Goucher: Phys. Rev. **81**, 475 (1951).
52. K. Seiler, H. Wucherer: N.T.F. **1**, 3 (1955).
53. E. M. Pell: Phys. Rev. **90**, 278 (1953).
54. E. L. Steele: J. Appl. Phys. **25**, 916 (1954).
55. We-Hsiung Ko: Sol. State Electronics **3,1** 59 (1961).
56. Leo Esaki: Phys. Rev. **109**, 603 (1958).
57. H. S. Sommers jr.: Proc. IRE **47**, 1201 (1958).
58. A. G. Chynoweth, W. L. Feldmann, R. A. Logan: Phys. Rev. **121**, 684 (1961).
59. R. Gremmelmaier, H. J. Henkel: Z. Naturforsch. **14a**, 1072 (1959).
60. R. L. Batdorf, G. C. Dacey, R. L. Wallace, D. J. Walsh: J. Appl. Phys. **31**, 613 (1960).
61. H. P. Kleinknecht: Sol. State Electronics **2,2/3**, 133 (1961).
62. N. Holonyak jr., I. A. Lesk, R. N. Hall, J. J. Tiemann, H. Ehrenreich: Phys. Rev. Lett. **3**, 167 (1959).
63. A. G. Chynoweth, G. H. Wannier, R. A. Logan, D. E. Thomas: Phys. Rev. Lett. **5**, 57 (1960).
64. A. R. Calawa, R. H. Rediker, B. Lax, A. L. McWhorter: Phys. Rev. Lett. **5**, 55 (1960).
65. K. K. N. Chang: Proc. IRE **47**, 1268 (1959).
66. M. E. Hines: Bell Syst. Tech. J. **39**, 477 (1960).
67. R. F. Trambarulo, C. A. Burrus: Proc. IRE **48**, 1776 (1960).
68. R. F. Trambarulo: Proc. IRE **48**, 2022 (1960).

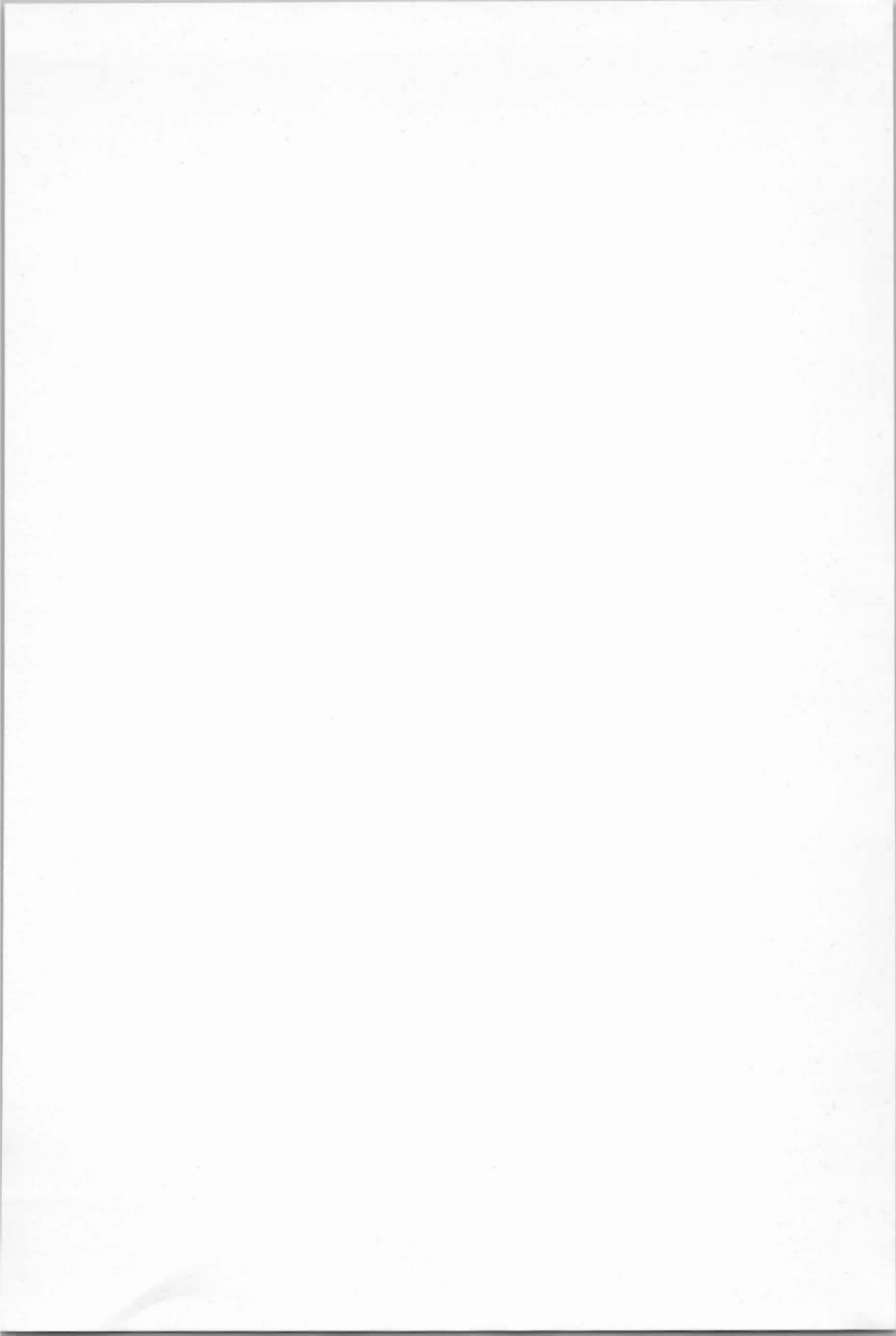
69. C. A. Burrus: J. Appl. Phys. **32**, 1031 (1961).
70. I. A. Lesk, H. A. Jensen: Sol. State Electronics **1,3**, 183 (1960).
71. H. J. Hartmann, M. Michelitsch, W. Steinhäuser: A.E.U. **15**, 125 (1961)
72. C. Zener: Proc. Roy. Soc. (London) **145**, 523 (1934).
73. K. B. McAffee, R. M. Ryder, W. Shockley, M. Sparks: Phys. Rev. **83**, 650 (1951).
74. K. G. McKay: Phys. Rev. **94**, 877 (1954).
75. D. R. Muss, R. F. Greene: J. Appl. Phys. **29**, 1534 (1958).
76. R. S. Claassen: J. Appl. Phys. **32**, 2372 (1961).
77. C. G. Thornton: Coloque International sur les dispositifs à Semiconducteurs, Band II, 725 (1961).
78. J. M. Early: Proc. IRE **40**, 1401 (1952).
79. A. R. Moore, J. I. Pankove: Proc. IRE **42**, 907 (1954).
80. J. Laplume: Compt. rend. **238**, 1107 und 1300 (1954).
81. A. J. Wahl: IRE Transactions **ED 4**, 216 (1957).
82. K. F. Stripp, A. R. Moore: Proc. IRE **43**, 856 (1955).
83. W. M. Webster: Proc. IRE **42**, 914 (1954).
84. H. Rüchardt: N.T.Z. **15,7**, 333 (1962).
85. C. A. Lee: Bell Syst. Tech. J. **35**, 23 (1956).
86. J. A. Hoerni: IRE Transactions **ED 8**, 178 (1961).
87. P. J. W. Jochems, O. W. Memelink, L. J. Tummers: Proc. IRE **46**, 1161 (1958).
88. H. C. Theurer, J. J. Kleimack, H. H. Loar, H. Christensen: Proc. IRE **48**, 1942 (1960).
89. S. Wang, T. T. Wu: IRE Transactions **ED 6**, 162 (1959).
90. R. Emeis, A. Herlet, E. Spenke: Proc. IRE **46**, 1220 (1958).
91. H. P. Kleinknecht: N.T.Z. **15,8**, 394, (1962).
92. L. D Armstrong, C. L. Carlson, M. Bentivegna: RCA Review **17**, 37 (1956).
93. L. J. Giacoletto: RCA Rev. **15**, 506 (1954).
94. R. Wiesner: N.T.Z. **15,7**, 323 (1962).
95. M. Ilschner, W. Steinhäuser: SEG-Nachrichten **1** (1957)

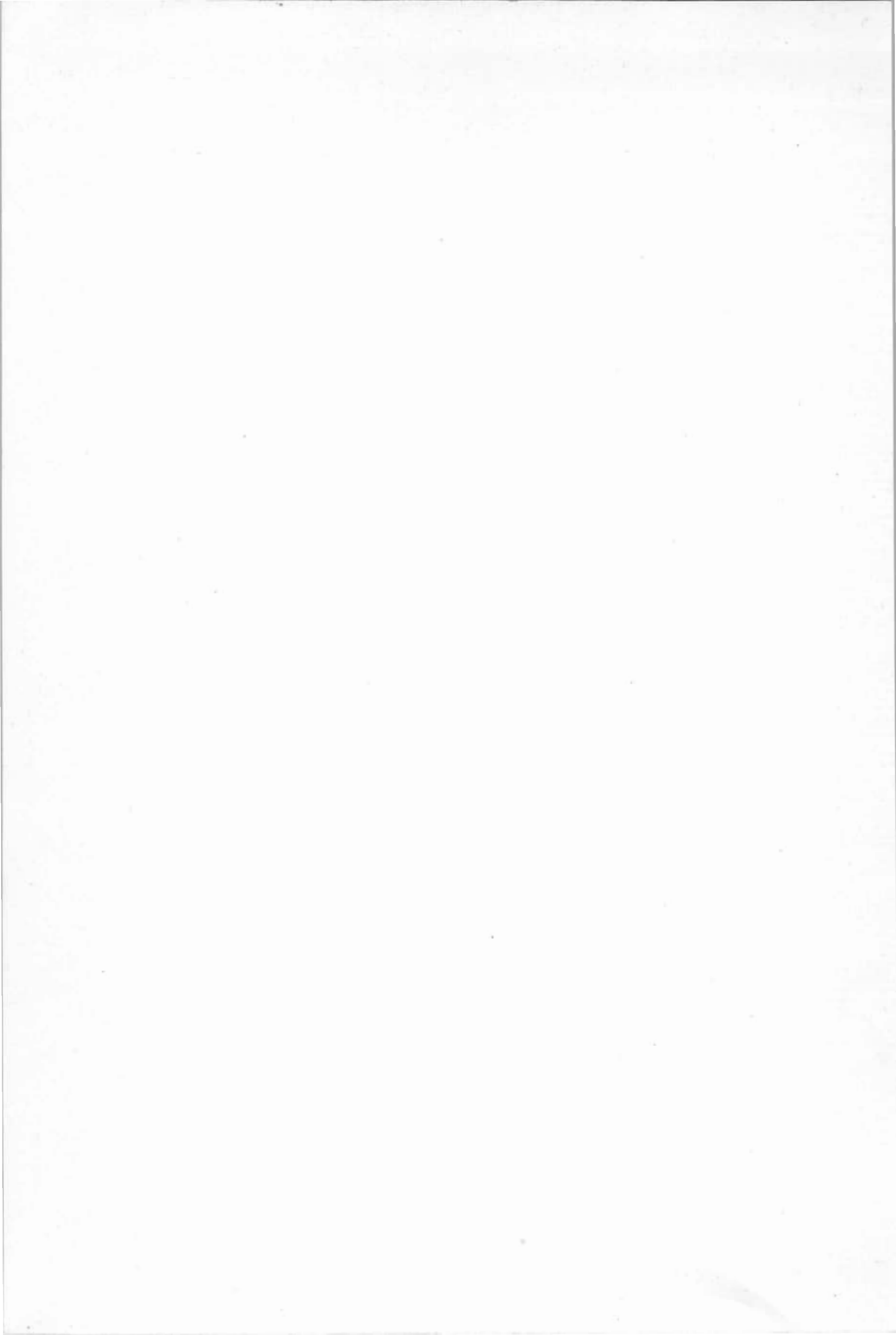
Namenverzeichnis

Sachverzeichnis









SEILER / PHYSIK UND TECHNIK DER HALBLEITER